

# **A network-based approach to estimating partisanship and analyzing change in polarization during the 2016 general election**

By

Michael W. Kearney

Submitted to the graduate degree program in the Department of Communication Studies  
and the Graduate Faculty of the University of Kansas in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy

---

Mary C. Banwart, Co-Chairperson

---

Jeffrey A. Hall, Co-Chairperson

Committee members

---

Benjamin R. Warner

---

Ashley Muddiman

---

Clayton M. Webb

Date defended:

---

20 April 2017

The Dissertation Committee for Michael W. Kearney certifies  
that this is the approved version of the following dissertation :

A network-based approach to estimating partisanship and analyzing change in polarization  
during the 2016 general election

---

Mary C. Banwart, Co-Chairperson

---

Jeffrey A. Hall, Co-Chairperson

Date approved:

20 April 2017

---

# Abstract

Communication and media research lacks an accessible and systematic approach to measuring political partisanship in decentralized media environments. In this dissertation, a network-based measurement of partisanship is proposed and then used to analyze social media users during a highly contentious general election. Study I (Chapter 2) introduces *rtweet*, a newly developed open-source software package designed to collect Twitter data. Study II (Chapter 3) then uses *rtweet* to gather publicly available Twitter data and demonstrate a network-based approach to estimating partisanship. Finally, Study 3 (Chapter 4) extends this network-based approach to analyze change over time in network polarization among partisan and non-partisan users during the 2016 general election. Results showcase the range and validity of network-based estimates of partisanship and provide clear evidence of partisan selective exposure and network polarization on Twitter as proximity to the election increases.

*Keywords:* networks, partisanship, polarization, selective exposure, twitter

## Acknowledgements

This project would have never gotten off the ground if it hadn't been for the love and support of a lot of people. It's hard for me to put into words how thankful and appreciative I am. And with how many times I've already written and erased these first few sentences, I imagine I may have an even harder time giving proper credit to all of the people without whom this dissertation wouldn't be. So, I've opted to keep things brief.

To my family, thank you for always being there for me, regardless of circumstance. And a special thank you to my mom, who has been my biggest cheerleader and supporter even though I'm sure she never imagined one person could spend this long in school.

To Casi, my best friend and future wife, thank you for being patient with me, looking out for me, [occasionally] challenging me, and, most of all, thank you for laughing with me.

To my advisors, Mary and Jeff, and the rest of my committee, thank you for pushing me, making me think harder, and, despite my own misgivings, showing unwavering confidence in me.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Concerns about new media . . . . .	2
1.2	Goals . . . . .	2
1.3	Overview . . . . .	3
<b>2</b>	<b>rtweet: Collecting Twitter data</b>	<b>4</b>
2.1	Contribution . . . . .	5
2.2	Authorization . . . . .	7
2.3	Examples . . . . .	8
2.3.1	Search tweets . . . . .	8
2.3.2	Stream tweets . . . . .	11
2.3.3	Get followers . . . . .	13
2.4	Resources . . . . .	16
<b>3</b>	<b>A network-based approach to estimating partisanship</b>	<b>18</b>
3.1	Estimating partisanship . . . . .	19
3.1.1	Partisan media fragmentation . . . . .	19
3.1.2	Problems with self report . . . . .	20
3.2	Defining partisanship . . . . .	21
3.2.1	Partisan networks . . . . .	22
3.2.2	Non-partisan networks . . . . .	23

3.3	Method . . . . .	24
3.3.1	Twitter data . . . . .	24
3.3.2	Source accounts . . . . .	24
3.3.3	Sample users . . . . .	25
3.3.4	Data wrangling . . . . .	27
3.4	Results . . . . .	29
3.4.1	Overview of tests . . . . .	29
3.4.2	Parallel analysis . . . . .	29
3.4.3	Rotated components . . . . .	29
3.4.4	Applied validity . . . . .	31
3.5	Discussion . . . . .	39
3.5.1	Measuring gradual change over time . . . . .	40
3.5.2	Using network-based approaches in other contexts . . . . .	40
3.5.3	Limitations . . . . .	41
<b>4</b>	<b>Analyzing change in network polarization</b>	<b>44</b>
4.1	Measuring polarization . . . . .	46
4.1.1	Voting records . . . . .	47
4.1.2	Issue positions and attitudes . . . . .	47
4.1.3	Media exposure . . . . .	48
4.2	Social networks . . . . .	49
4.2.1	Social media . . . . .	49
4.2.2	Partisan selective exposure . . . . .	50
4.2.3	Twitter . . . . .	51
4.2.4	Previous research . . . . .	53
4.3	Change in polarization . . . . .	54
4.3.1	Analysis of change . . . . .	54
4.3.2	Political salience and the election . . . . .	54

4.4	Method . . . . .	56
4.4.1	Data collection . . . . .	56
4.4.2	Sample users . . . . .	57
4.4.3	Follow decisions . . . . .	58
4.5	Results . . . . .	60
4.5.1	Multilevel modeling . . . . .	60
4.5.2	Models . . . . .	62
4.5.3	Interpreting the interaction . . . . .	62
4.6	Discussion . . . . .	68
4.6.1	Network polarization . . . . .	68
4.6.2	Limitations . . . . .	71
<b>5</b>	<b>Conclusion</b>	<b>73</b>
5.1	Summary . . . . .	74
5.1.1	Technological advancements . . . . .	74
5.1.2	Estimating partisanship and analyzing change . . . . .	74
5.1.3	Extending the network-based approach . . . . .	75
5.2	Communication, the hard science . . . . .	75
5.2.1	Computational social science . . . . .	75
5.2.2	Future directions . . . . .	76
<b>6</b>	<b>Appendices</b>	<b>78</b>
6.1	Study II . . . . .	79
6.1.1	User groups . . . . .	79
6.1.2	Ordinary least squares model . . . . .	80
6.1.3	Negative binomial models . . . . .	83
6.2	Study III . . . . .	84
6.2.1	Missing data . . . . .	84

6.2.2	Source accounts . . . . .	85
-------	---------------------------	----



# List of Figures

2.1	Plotting multiple time series of tweets frequency . . . . .	12
3.1	Eigenvalues of follow decisions by users sampled from twelve source accounts	30
3.2	Rotated estimates made by users sampled from twelve source accounts . . .	32
3.3	Rotated component scores of elite accounts . . . . .	35
3.4	Estimate of partisan orientation of each elite's network . . . . .	36
3.5	Rotated component scores of elite accounts . . . . .	37
3.6	Comparing estimates of follow decisions targeting congressional elites with estimates of follow decisions targeting elites determined by a network-based approach . . . . .	38
4.1	Individual trajectories . . . . .	64
4.2	Change in number of weighted partisan follow-decisions by partisan versus non-partisan group . . . . .	65
4.3	Change in number of weighted partisan follow-decisions by partisan groups .	66
4.4	Change in number of partisan follow-decisions by all groups . . . . .	67
6.1	Twitter Source Accounts . . . . .	79
6.2	Change in partisan scores by account . . . . .	86

# List of Tables

3.1	Descriptive statistics of source accounts . . . . .	26
3.2	User summary statistics by source accounts . . . . .	27
3.3	Standardized rotated estimates of follow decisions made by users sampled from twelve source accounts . . . . .	31
3.4	Rotated component correlations . . . . .	33
3.5	Mean number of follow decisions of congressional and Twitter elites by parti- san sample group . . . . .	34
4.1	Summary statistics of sample and group . . . . .	58
4.2	Varying intercept/slope estimates predicting change in homogeneous follow decisions . . . . .	61
6.1	OLS model estimates predicting change in partisan follow decisions . . . . .	81
6.2	Negative binomial estimates predicting change in network polarization . . . . .	83
6.3	Number of observations per time point by group . . . . .	84

# Chapter 1

## Introduction

## 1.1 Concerns about new media

People often want to know whether new media is good or bad. They may pose the question in a number of different ways, for example, “Does new media increase or decrease productivity?”, “Does new media help or hurt democracy?”. But the underlying question is the same one that has been asked and answered a countless number of times in human history. It is the same question that was asked shortly after the invention of the printing press and then again after people started gathering in the same room to listen to a radio show broadcast across the country. It is the same question that was asked when motion pictures were projected at the cinema, when television shows were transmitted across the country, and when hundreds of digital cable channels were displayed via funny-looking set-top boxes. It is the same question that was asked of cell phones, internet chat-rooms, instant messengers, and videos on-demand. It is even the same question that persists despite the entire world became digitally connected in a moment of Pokémon Go. And it will continue to be the same question for as long as we have advancements in media and ample opportunity in space-time.

## 1.2 Goals

The aim of this project is not to satisfy the demands of an age-old expression of moral panic, but to explore the possibilities introduced by new media technologies. In particular, the goals of this project are two-fold: 1. to provide a meaningful contribution to our theoretical understanding of partisan selective exposure and political polarization as manifested in the current media environment, and 2. to demonstrate a new approach to studying and measuring social science phenomena that takes advantage of the changing media environment in order to improve the perspective and maximize the reach of communication research.

## 1.3 Overview

Currently missing from communication and media research is an accessible, systematic, and coherent approach to measuring political partisanship in decentralized media environments. In this dissertation, a network-based measurement of partisanship is proposed and then used to analyze social media users during a highly contentious general election. The chapters that follow describe, use, and then extend this method in different ways. In Chapter 2 (Study I), a newly developed open-source software package, *rtweet*, is introduced as a tool designed to collect Twitter data. Chapter 3 (Study II) then uses *rtweet* to gather publicly available Twitter data and demonstrate a network-based approach to estimating partisanship. Finally, Chapter 4 (study III) extends this network-based approach to estimating partisanship by analyzing change in network polarization over time among partisan and non-partisan users during the 2016 general election. Results showcase the range and validity of network-based estimates of partisanship and provide clear evidence of partisan selective exposure and network polarization on Twitter as proximity to the election increases. Finally, Chapter 5 reflects on the overall project and considers future directions of research.

# Chapter 2

## rtweet: Collecting Twitter data

*Abstract:* Interest in Twitter data continues to grow, but for many interacting with Twitter's API seems like a daunting task. With the deprecation of *twitteR*, R users can now turn to the new *rtweet* package. An introduction to *rtweet* is provided, including step-by-step instructions for authorization methods and examples of several of *rtweet*'s major functions, including `lookup_statuses`, `search_tweets`, `stream_tweets`, `get_followers`, `lookup_users`, and `get_timeline`.

*Keywords:* partisanship, selective exposure, social networks, twitter

Social media activity produces massive amounts of user data every day. Social media companies often make this data available to the public through an application program interface (API). Web APIs are protocols and procedures implemented in order to regulate user interactions with a site. In theory, one could endlessly [web] scrape unstructured social media data by rendering HTML pages and then crawling each page. But this practice can lead to excessive traffic and strain on a server. Plus, it violates most sites’ terms of service. Providing a public API, on the other hand, gives firms a way to regulate data requests and, at the same time, encourage development of platform-specific applications and integration with other non-native technologies. Twitter, for example, offers two major public APIs. However, despite [excellent documentation](#), for many R users, interacting with Twitter’s public APIs—creating and sending API requests, reading in the data from the site, balancing requests with Twitter rate limits, and, once the data are actually transferred, organizing the data—remains a daunting task.

Twitter provides two major public APIs, a [REST API](#) and a [stream API](#). The REST API returns descriptive information on users, statuses posted to user timelines, and search results of statuses<sup>1</sup> posted in the previous 6-9 days. The stream API returns a live feed of randomly sampled or filtered statuses generated by users in real-time. Both the REST and stream API are available to anyone with a developer account.

## 2.1 Contribution

To date the most popular R package for interacting with Twitter’s APIs has been the *twitteR* package ([Gentry, 2013](#)). The package is flexible and widely used (e.g., [Barberá & Rivero, 2014](#); [Breen, 2012](#); [Harris, 2015](#); [Robinson, 2016](#); [Vidal et al., 2015](#); [Younis, 2015](#)), but maintenance of *twitteR* has faded over time, and after years of reliable service, the package has been deprecated ([Gentry, 2013](#)), leaving R without an up-to-date API wrapper package

---

<sup>1</sup> To be consistent with Twitter documentation, the word “statuses” will be used to refer to Twitter posts submitted by users, or what most users commonly refer to as “tweets”

for collecting Twitter data. As of August 2016, *twitteR* users have been [encouraged to transition to \*rtweet\*](#). In the short term, this means issues with *twitteR* are unlikely to be resolved, whereas *rtweet* will have more active support. Reasons for transitioning from *twitteR* to *rtweet* will continue accumulating as Twitter [continues to implement changes to its APIs](#). At least one recent change to Twitter’s REST API, the inclusion of an “extended text mode,” is worth noting, as it will soon render the compatibility mode on which *twitteR* is built obsolete.

The *rtweet* package provides a few extra features and new conveniences in addition to providing the same basic functionality as the *twitteR* package. As for conveniences, users may find *rtweet*’s distinct treatment functions versus data more approachable than *twitteR*’s use of the reference class. In contrast, *rtweet* returns, as its default, data frames. Another convenience introduced in *rtweet* has to do with the ease and documentation for setting up user access tokens. Anyone who has ever tried to gain authorization to access Twitter’s APIs will likely appreciate this contribution. In the section that follows, users will be guided through the authorization process. Code is then provided to create an environment variable, which should save users from having to do this everytime they want to work with Twitter’s APIs. Before getting to the authorization methods and code examples, however, the motivations behind developing *rtweet* are described below.

The goals of *rtweet* are (a) to make interacting with Twitter’s APIs more approachable for less-experienced users and (b) to make the structure and format of Twitter data more accessible to researchers. Development of *rtweet* to date has mostly focused on objectives of data collection and research.<sup>2</sup> Regardless of goals, however, feedback and suggestions are encouraged. Issues are welcome on Github, provided users include all relevant code along with error message and session info. The best place to [create a new issue is on Github](#). For other issues and discussion, users are welcome to [join and post to the Google group](#).

---

<sup>2</sup> Although, currently, *rtweet* offers users several functions for posting on behalf of their own Twitter accounts, but efforts to add documentation and support for those functions has nevertheless lagged



## 2.2 Authorization

One of the biggest barriers preventing people from working with Twitter's APIs is the authorization process. In order to have authorized access to Twitter's APIs, users must first create a Twitter application. For the software-averse, this is no easy task. The process is therefore broken down step by step below. Included in these steps are procedures for saving authorization information as an environment variable, which means users only have to go through this process once.

- Navigate to <https://apps.twitter.com> and create a new application by providing a name for the application, brief description, and a website of your choosing (placeholder URLs are acceptable). In the Callback URL field, make sure to enter the following:  
`http://127.0.0.1:1410`
- Check yes if you agree and then click "Create your Twitter application"
- Once you've successfully created an application, click the tab labeled Keys and Access Tokens to retrieve your consumer and secret keys.
- To create a personal access token, users must provide their Consumer Key (also referred to as "API Key") and Consumer Secret (also referred to as "API Secret").

```
## name assigned to created application
app <- "appname"

## api key (demonstration only)
consumer_key <- "a4aoi3uaysf9as79d623409as"

## api secret (demonstration only)
consumer_secret <- "a89s79dfkl151134098akadskfj5623vbfewfw7adks11234098"

## create token
```

```
token <- create_token(app, consumer_key, consumer_secret)
```

If this is the first time authorizing the application, a browser should pop-up with an option for approving use of the application. Once authorized, it is easiest to save the token to the user's home directory and then store its path to the user's .Renviron file. This process can be handled with a few lines of code, which are provided below:

```
## home directory
home <- path.expand("~/")

## save token
saveRDS(token, file = file.path(home, ".rtweet_token"))

## save path to token as environment variable
cat(paste0("TWITTER_PAT=", file.path(home, ".rtweet_token")),
    file = file.path(home, ".Renviron"),
    fill = TRUE, append = TRUE)

## retrieve token
get_tokens()
```

## 2.3 Examples

### 2.3.1 Search tweets

Perhaps the most popular feature offered by Twitter is their search API. Users can search for a wide variety of tweets from across the world. Searches can even specify specific geographical locations and user languages.

```
rt <- search_tweets(q = "", lang = "pt",
                    geocode = "-22.912214,-43.230182,1km")
```

Alternatively, users can access the stream API to find the most popular tweets from a user that include a certain hashtag.

```
cmdr <- search_tweets("from:@Cmdr_Hadfield #NASA")
```

Unfortunately, data returned by `search_tweets()` is limited to the most recent 6-9 days. Through the public API, there is no way to search all Tweets. In the following example, *rtweet* is used to collect data on tweets sent by Donald Trump's Twitter account. Although `get_timeline` is limited to the most recent 3,200 tweets, it is possible to get all tweets associated with a user if you have the status Id's associated with those tweets. The status Id's of tweets sent from `realDonaldTrump`'s account used in the following examples can be downloaded at [https://github.com/mkearney/rtweet/blob/master/data/trump\\_ids.rds](https://github.com/mkearney/rtweet/blob/master/data/trump_ids.rds). After reading in the Id's, use `lookup_statuses` to retrieve tweets data for each ID. The first six rows of the returned data frame are printed below.

```
## read status Id's of Trump tweets
ids <- readRDS(paste0("https://github.com/mkearney/rtweet/blob/",
                     "master/data/trump_ids.rds?raw=true"))

## lookup statuses (tweets) data
djt <- lookup_statuses(ids)

## preview data
head(djt)
```

```

##      screen_name  user_id      created_at  status_id
## 1realDonaldTrump 25073877 2009-05-20 22:29:47 1864367186
## 2realDonaldTrump 25073877 2010-05-17 16:48:39 14171725850
## 3realDonaldTrump 25073877 2010-03-23 16:26:23 10932826995
## 4realDonaldTrump 25073877 2010-04-13 13:49:15 12104502409
## 5realDonaldTrump 25073877 2010-02-26 21:18:26 9696989458
## 6realDonaldTrump 25073877 2010-03-15 18:21:52 10528960360
##
##                                     text
## 1 Read a great interview with Donald Trump that appeared..
## 2 Congratulations to our new Miss USA, the beautiful Rima.
## 3 The Trump Hotel Collection is currently nominated for...
## 4 Looking forward to seeing the World Champion Yankees....
## 5 From Donald Trump: \"I'm so proud of my wife Melania....
## 6 The new season of the Celebrity Apprentice is off to a..
##  retweet_count favorite_count is_quote_status is_retweet
## 1              6              5             FALSE      FALSE
## 2             115             186             FALSE      FALSE
## 3              13              6             FALSE      FALSE
## 4              55              21             FALSE      FALSE
## 5              25              22             FALSE      FALSE
## 6              23              6             FALSE      FALSE
##
##      lang      source
## 1      en  Twitter Web Client
## 2      en  Twitter Web Client
## 3      en  Twitter Web Client
## 4      en  Twitter Web Client
## 5      en  Twitter Web Client
## 6      en  Twitter Web Client

```

In addition to data retrieval and parsing functions, the *rtweet* package provides some convenience plotting functions as well. Using `ts_filter` it is possible to transform tweets data into time-series-like frequency data. If regular expressions are passed to the `filter` argument, the function will return multiple time series, one related to each of the supplied filters. In tidy form, the data are ready to be handed off and plotted by `ggplot2`.

```

## filter tweets frequency using regular expressions
dts <- ts_filter(djt, by = "2.5 months",
                filter = c("sad", "dishonest", "week", "failing"))

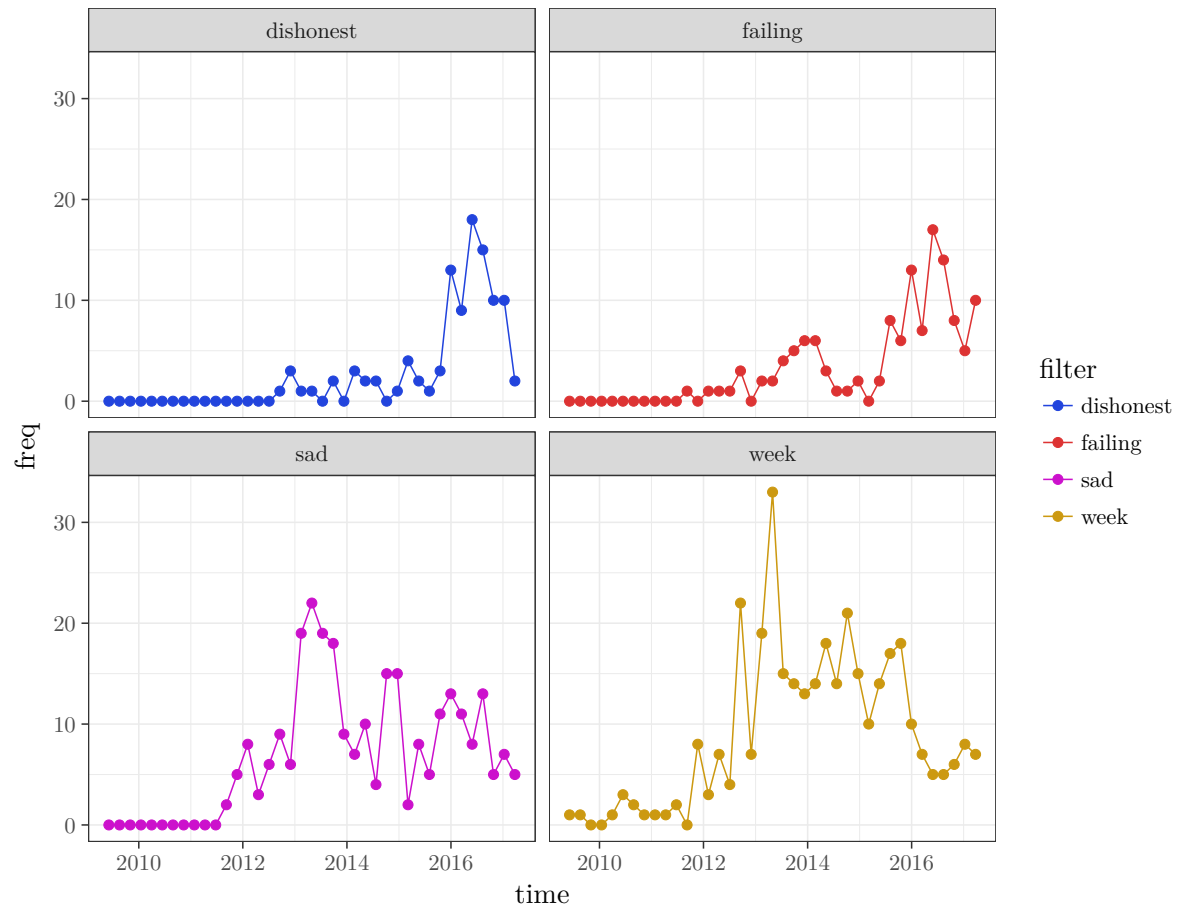
## plot with ggplot2
library(ggplot2)
ggplot(dts, aes(x = time, y = freq, color = filter)) +
  geom_line() +
  geom_point(size = 1.5) +
scale_color_manual(
  values = c("#2244dd", "#dd3333", "#cc11cc", "#cc9911")) +
  theme_bw() +
facet_wrap(~filter, ncol = 2)

```

### 2.3.2 Stream tweets

In addition to interacting with the REST API, users can also use *rtweet* to access the stream API. Because the public search API is limited to only the previous 6-9 days, many users will maximize their returns by using the stream API to capture upcoming or active events. Using the code provided below, for example, users can collect all public tweets mentioning *rstats* or *crantastic* for 10 minutes. In addition to returning a data file, the **file\_name** argument also allows users to save the json file to a location of their choosing. The stream is written straight into the json file, so for larger streams users may consider adding the argument **parse = FALSE**, as parsing large files may run into limits of system resources. Saved json files can be parsed at any time using the **parse\_stream** function, which has the added benefit of parsing the file in pieces.

Figure 2.1: Plotting multiple time series of tweets frequency



```
## set timer (in seconds) for 10 minutes
streamtime <- 60 * 10

## provide filter keywords
q <- "rstats,crantastic"

## initiate stream
rt <- stream_tweets(q = q,
                    timeout = streamtime,
                    parse = FALSE,
                    file_name = "livestream.json")
```

### 2.3.3 Get followers

Using the Twitter’s Rest API, one can specify a user and request a list of all the Twitter Id’s of users who follow that specified user. So, if one wanted to gather data on a random sample of users who follow **ProPublica**, then using Twitter’s REST API one could request user Id’s of all followers. Due to Twitter’s API rate limits, users can only retrieve a maximum of 75,000 follower Id’s every 15 minutes. Using a for loop, this process can still be automated to return millions of users. User ids returned from **get\_followers** could then be passed along to **get\_friends**, which would then return the friend networks for each user—that is, all the accounts a user chooses follow. If one were to do this for all ten thousand users retrieved using **get\_followers**, then each of those ten thousand friend networks should include the original user from which the followers were identified—in the example used earlier in this paragraph, that would be **ProPublica**—in addition to the, presumably, hundreds of thousand, or even millions, of other accounts found in the friend networks.

If the user has more than 75 thousand followers (the maximum allowed per token per 15 minutes), then automating the collection of all of a user’s followers requires some extra code. In the demonstration below, a “for loop” is used to collect the roughly 650,000 followers of **ProPublica**. The loop is setup by first setting **N** equal to the total number of followers (650,000) divided by 75,000 (75,000), the maximum number of user Id’s retrieved by a single function call. The code below makes this calculation and applies **ceiling** such that the output is rounded UP to nearest integer.

```
N <- ceiling(650000 / 75000)
```

The stored **N** value is then used to create an empty list vector, which will store the output from iterations of the loop.

```
fds <- vector("list", N)
```

Next, a `page` object is created and initially set to `NULL`. Each successive function call will return, along with a data frame of user Id, a unique cursor (a string scalar), which can be accessed with `next_cursor`. Thus, in every loop after the first, `page` should be a non-null value that communicates to Twitter's API where the loop last left off.

```
page <- NULL
```

Running the “for loop” without any kind of pause or time-delaying mechanism will quickly run into API rate limits. In the demonstration code provided below, the base R function `Sys.sleep` is used to wait for the rate limit to reset between iterations.

```
## run loop N times
for (i in seq_len(N)) {
  ## get 75,000 followers at a time
  fds[[i]] <- get_followers("ProPublica", page = page)
  ## store next cursor value in page
  page <- next_cursor(fds[[i]])
  ## print message for verbose output
  message(paste(i, "of", N, "loops complete!", sep = " "))
  ## if this is the Nth loop, then it's done!
  if (i == N) break
  ## sleep 15 minutes for rate limit reset
  Sys.sleep(60 * 15)
}
```



```
}
```

Once the loop is complete, the first line of code below takes the list vector of data frames and collapsed it down into a single data frame. The first six rows of are then printed using the `head` function.

```
## collapse Id into single data frame
fdsdat <- do.call("rbind", fds)
## preview data
head(fdsdat)
```

```
## # A tibble: 6 x 1
##       user_id
##       <dbl>
## 1    249833800
## 2    2201989099
## 3    2354361330
## 4    376743481
## 5    498985509
## 6  818654555588149248
```

In addition to user Id, it is also possible to get a variety of users-specific meta data. For instance, same data frame of user Id can be passed along to `lookup_users`. Again, however, users will run into Twitter rate limits, which in this case is a hardly-limiting 90 thousand users per 15 minutes. In theory, it would be possible to again loop though all of the users. But data returned by `lookup_users` is a lot larger than `get_followers`. In many cases, taking a random sample of a population of users may be a desirable middle ground. This is again demonstrated in the code below.

```
## desired sample size. max per token is 90000 (9e4)
N_sample <- 9e4

## randomly sample user IDS
sampids <- sample(unique(fdsdat[["user_id"]]), N_sample)

## lookup users data
sampusrs <- lookup_users(sampids)
```

User Id may also be passed to `get_timeline`, which allows retrieval of up to 3,200 of the most recent statuses (tweets) posted by a user.

```
## get timeline (tweets data)
usr1 <- get_timeline(sampusrs[["user_id"]])
```

## 2.4 Resources

For additional resources on *rtweet*, visit the package documentation website at the following URL: <https://mkearney.github.io/rtweet>. For a list of blog posts featuring the *rtweet* package, see the following:

- <https://r-bloggers.com/faces-of-rstats-twitter>
- <https://r-bloggers.com/a-glance-at-r-bloggers-twitter-feed>
- <https://r-bloggers.com/twitter-coverage-of-the-lorne-genome-conference-2017>
- <https://r-bloggers.com/the-animals-of-actuallivingscientists>
- <https://r-bloggers.com/analyzing-first7jobs-tweets-with-monkeylearn-and-r>

- <https://r-bloggers.com/surveillance-out-of-the-box-the-zombie-experiment>
- <https://citizensciences.net/2017/01/26/4-things-twitter-tells-us-about-citizen-science>
- <https://livefreeordichotomize.com/2017/03/16/enar-in-words>
- <https://mytinyshinys.com/post/twitterFollowersCollage>

# Chapter 3

## A network-based approach to estimating partisanship

*Abstract:* Communication and media research lacks an accessible and systematic approach to measuring political partisanship in decentralized media environments. The current study demonstrates a network-based approach to estimating partisanship, using publicly available Twitter data and open-source software. Results reveal the range and validity of network-based estimates of partisanship and provide clear evidence of partisan selective exposure in user networks on Twitter. Consideration is also given to alternative uses of the demonstrated approach.

*Keywords:* partisanship, selective exposure, social networks, twitter

## 3.1 Estimating partisanship

### 3.1.1 Partisan media fragmentation

The rise of media choice has made it difficult to measure political partisanship of a diverse and wide range of media sources. Subsequently, our ability to study and understand selective exposure in decentralized media environments remains in question. Many approaches to estimating partisanship found in the literature rely on data collected via self-report (e.g., [Baldassarri & Gelman, 2008](#); [DiMaggio et al., 1996](#); [Hmielowski et al., 2015](#); [Iyengar et al., 2008](#); [Lau et al., 2016](#); [Lee et al., 2014](#); [Lee, 2016](#); [Levendusky & Malhotra, 2016](#); [Levendusky, 2013](#); [Stroud, 2010](#)). These approaches measure the partisanship of media sources, such as MSNBC or Fox News, by asking respondents about their voting history, attitudes toward political parties, and exposure to different kinds of media content. To make inferences from self-report data like this, however, researchers must then assume that respondents are fully aware of their media-selection behaviors. In today’s fast-paced, information-rich, and technology-filled media environment, this assumption is, at best, tenuous ([Knobloch-Westerwick & Johnson, 2014](#)). People continuously process new information. Given a finite amount of cognitive effort, homeostatic informational needs, and a media environment saturated with information, it is hard to imagine people are fully aware during every media-selection decision.

The problems related to self-report data in media exposure research are well known. A lesser acknowledged but equally concerning problem relates to the definition of [political] partisanship. Broadly, partisanship refers to political affiliations, or associations, in relation to organized groups. In other words, partisanship describes the distribution of units—voters, elected officials, media entities, etc.—relative to political parties. Despite similarities to social networks, however, current approaches tend to define partisanship using attitudes or media content and not, strictly speaking, networks. In a fragmented media environment, methods used to estimate partisanship must be broad enough to cover specialized and fleeting media sources while flexible enough to detect and account for changes in political affiliations.

Analysis that is limited to user attitudes or media content seems unlikely to satisfy these demands. Hence, data describing the connections between users on social media platforms, such as Facebook or Twitter, which are increasingly made accessible by open-source software packages, provide researchers with a treasure trove of publicly available and reliable data that directly captures political affiliations of media users and media sources.

The purpose of the present study is to develop a new method which enables a network-based approach to measuring partisanship. This method will be demonstrated using Twitter data collected with *rtweet*, an R package for collecting Twitter data (Kearney, 2016). This newly developed method will more accurately reflect the affiliative nature of partisanship while also avoiding methodological limitations of self-report. The primary contributions of this investigation are two fold. First, this study provides an empirical account of the extent to which political partisanship explains the organization of user networks on a popular social media platform. Second, the current investigation provides researchers with a network-based approach to measuring the partisanship of media sources. In conclusion, this method will enable researchers to estimate partisanship of a wide range of media sources using an expansive and publicly accessible source of real-time data.

### 3.1.2 Problems with self report

Setting aside the tenuous assumption that users are completely aware of their media selection behaviors, self-report based measurements of partisanship suffer from poor respondent recall because even if people are fully aware at-the-time of media exposure decisions, their ability to accurately recall those decisions later on, is notoriously unreliable (Clay et al., 2013; Prior, 2009a,b). Innovative research designs may reduce the tendency of individuals to over-report overall exposure to media (e.g., Prior, 2009b), but even the most sophisticated survey-based methods are unlikely to eliminate all sources of bias. And while advances in experimental designs show some promise (e.g., Arceneaux et al., 2012, 2013), active media choices, like those induced by experimental stimuli, are more likely to include attitude-discrepant

information than passive media choices (Smith et al., 2008). And even if researchers could somehow address all of the limitations in the measurement of media exposure, there would be no way to verify the validity of the data without some kind of passive media tracking technology. And, to date, few researchers have access, let alone resources, to use passive tracking devices.

## 3.2 Defining partisanship

Partisanship refers to the degree to which one affiliates or associates with a political party (Kenski, 1980). In other words, partisanship describes the degree to which one is political—in an organized, networked sense—and partisan orientation describes the political group with which one is affiliated. Partisanship differs from political ideology because it describes attitudes and behaviors toward organized parties and not political philosophies. When the two conflict, partisan motivations have even been found to supersede commitments to political ideology (Glaeser & Sunstein, 2013; Kahan et al., 2013; Kahneman, 2011; Warner & McKinney, 2013). As such, partisanship is not static. It is a snapshot reflection of competing political parties.

Because they are not logically constrained by ideological commitments, partisan platforms are free to fluctuate and often do between election cycles. Given its tumultuous and somewhat arbitrary state of existence, scholars have struggled to find reliable ways to estimate partisanship. Using publicly available voting records and party-nominations, researchers have developed numerous ways to measure partisan orientations of elected officials. But without obvious and publicly available indicators of partisanship, far fewer attempts have been made to measure partisan orientations of voters, non-voting public figures, and media entities.

### 3.2.1 Partisan networks

At its core, partisanship is driven by political affiliations. So while a political party may unite around ideologies or opinions, it is associations between people and entities, and not the particular ideas, that make the group partisan. Ultimately, then, partisanship is an expression of positive associations among a group of people and/or entities. The affiliative nature of partisanship is perhaps most obvious in the context of social media, where user networks consist of explicitly defined connections between users and, in the case of Twitter, well known media sources. This explains why, on social media for example, the presence of social endorsements, which highlight affiliative connections, influence media selection decisions regardless of the perceived partisan alignment of the media’s content ([Messing & Westwood, 2012](#)).

The rise of Twitter and increasingly available digital sources of “big data” provide researchers with new and unique opportunities to investigate political partisanship at the network-level. Although researchers have had relatively little time to study social media, a few studies give us some idea of what to expect on Twitter. [Himelboim \(2014\)](#) found Twitter conversation networks of political topics occurred largely within partisan clusters. [Barberá \(2015\)](#) went a step further, using *follow-decisions*, or whether a user decides to follow elites, to estimate the political ideology of [political] elite and mass public users. When limited to a “number of target users that includes politicians, think tanks, and news outlets with a clear ideological profile that span the full range of the ideological spectrum” (p. 80), [Barberá \(2015\)](#) demonstrates that follow-decisions serve as a useful tool in estimating partisan orientations of elected officials and mass public users.

Results from these approaches to measuring political ideology on Twitter vary depending on the distribution and connections of users. In theory, partisan network homophily could manifest in different ways, depending on the environment and the salience of different characteristics of network connections. With that said, one would reasonably expect the basic structural pattern of two-party partisanship to hold. In other words, republican users should



follow more republican-oriented accounts and democrat users should follow more democrat-oriented accounts. If partisan network composition on Twitter can be explained via partisan orientations, then the average follower of a well-known republican elites should follow other republican-leaning political elites. Likewise, the average follower of a well-known democratic elite should follow other democratic-leaning political elites.

### **3.2.2 Non-partisan networks**

The political consequences of increased media choice brought on by the digital age have been linked by many to perceived increases in political polarization, or the degree to which competing partisan affiliations diverge ([Prior, 2013](#)). Although evidence of polarization among the political elite is clear, some research challenges whether polarization has occurred in the general public (e.g., [Hetherington, 2001](#)). An increasingly popular theory regarding polarization of the general public is that partisans appear more divergent in recent years because non-partisans, or moderates, are essentially “tuning out” of politics by selecting more entertainment media options ([Layman & Carsey, 2002](#); [Levendusky, 2013](#), e.g.,). Given a vast pool of entertainment accounts active on Twitter, it seems reasonable to expect the role of partisanship to be less visible in the follow decisions of non-partisan users, or users who may otherwise be considered “moderate” compared to follow decisions made by partisan users.

## 3.3 Method

### 3.3.1 Twitter data

Twitter data were collected in June of 2016 during a highly contested and widely covered general election, shortly after U.S. presidential candidates from both major parties—Hillary Clinton for the Democrats and Donald Trump for the Republicans—had functionally secured the nomination for their respective political party.<sup>1</sup> Twitter makes large amounts of data available to the public via Application Program Interfaces (APIs). APIs provide sets of routines that describe how applications should be used. The current study access Twitter’s REST API,<sup>2</sup> which enables data collection of user networks. A free, open-source R package, *rtweet* (Kearney, 2016) was used to identify followers of well-known partisan and non-partisan source accounts and then to record friend networks of those identified followers. The *rtweet* package formulates HTTP requests, communicates with Twitter’s APIs, retrieves and wrangles response objects into tabular structures, and returns data to the user. Applications such as this have made it easier than ever before for researchers to systematically collect and analyze real-time Twitter data.<sup>3</sup>

### 3.3.2 Source accounts

The population for the current study consists of all followers of four well-known partisan media accounts for each partisan group and four well-known entertainment media accounts for the entertainment group. To ensure the selected source accounts were popular enough to attract large numbers of intended users but not so popular that they transcended partisan

---

<sup>1</sup> Donald Trump clinched the Republican nomination on May 4, 2016. Hillary Clinton clinched the nomination on June 6, 2017. Data were collected on June 13, 2017.

<sup>2</sup> Twitter offers multiple public APIs, but two in particular—REST (archive) and stream (live)—allow users to harness data generated by Twitter users. The current study uses Twitter’s REST API because includes routines for retrieving data on user-level networks (friends and followers).

<sup>3</sup> For a similar library in the python environment, see Tweepy.

considerations,<sup>4</sup> accounts were considered if they had at least one-hundred thousand followers and fewer than five million followers. Accounts were also considered if they were used in previous research. For example, studies routinely reference Fox News, Sean Hannity, Bill O’Reilly, and Glen Beck for republicans and Huffington Post, Rachel Maddow, Jon Stewart, and Keith Olberman for democrats (e.g., [Arceneaux et al., 2012, 2013](#); [Holbert et al., 2012](#); [Wicks et al., 2014](#)).

Source accounts selected to represent the partisan republican group included the Drudge Report (@DRUDGE\_REPORT), Fox News Politics (@foxnewspolitics), Sarah Palin (@SarahPalinUSA), and Sean Hannity (@seanhannity). Source accounts selected to represent the partisan democrat group included Huffington Post Politics (@HuffPostPol;  $N = 170,000$ ), Rachel Maddow (@maddow;  $N = 350,000$ ), Paul Krugman (@paulkrugman;  $N = 375,000$ ), and Salon.com (@Salon). Lastly, source accounts representing the non-partisan entertainment group included AMC TV (@AMC\_TV), American Idol (@AmericanIdol), Sports Illustrated (@SInow), and CBS’s Survivor (@survivorCBS). Descriptive information for the source accounts can be found in Table 3.1.

### 3.3.3 Sample users

Unique identifiers, or *user Id’s*, were collected for all followers of each of the selected source accounts using the *rtweet* function `get_followers`. Once all followers were identified, user Id’s were pooled into the three groups—the presumed groups associated with each source account—and a number of users, which was more conducive to API rate limits, was randomly selected from each group ( $N = 60,000$ ; 20,000 users per group). To filter out spam, bot, and inactive accounts, user-level [meta] data was retrieved using *rtweet*’s `lookup_users()` function. Guided by first-hand experience and previous research (e.g., [Barberá, 2015](#); [Haustein et al., 2016](#); [Yardi et al., 2009](#)), several filters were selected and then applied to user-level data. Users were removed if they were *protected* (set their accounts to private), had fewer

---

<sup>4</sup> For instance, at the time of writing, President Obama had 75 million followers, meaning his popularity likely watered down the influence of partisan affiliations.

Group	Screen Name	Followers	Friends	Created	Tweets
Dem	maddow	5,117,568	2,385	2008-09-04	4,311
Dem	Salon	847,106	6,046	2008-10-24	130,395
Dem	HuffPostPol	904,212	9,052	2008-07-16	257,501
Dem	paulkrugman	1,697,264	5	2008-10-27	7,388
GOP	seanhannity	1,609,911	6,711	2009-05-21	32,358
GOP	SarahPalinUSA	1,285,051	137	2009-08-13	3,144
GOP	DRUDGE_REPORT	1,006,165	2	2008-05-06	175,875
GOP	foxnewspolitics	650,361	247	2008-08-28	39,396
Ent	AMC_TV	586,398	220	2010-05-27	11,370
Ent	AmericanIdol	2,085,964	5,785	2009-08-05	75,411
Ent	SInow	1,559,428	661	2009-04-02	175,420
Ent	survivorcbs	483,600	131	2009-08-27	8,981

Table 3.1: Descriptive statistics of source accounts

than 50 or greater than 1,500 friends or followers, fewer than 200 tweets, or failed to post a status an average of once every ten days.<sup>5</sup> Finally, to ensure consistency and to make navigating Twitter’s API rate limits more reasonable, an equal number of users ( $n = 1,000$ ) were sampled for each group, yielding a final sample size of 3,000 users ( $N = 3,000$ ). User summary statistics for each source account can be seen in Table 3.2. Unfortunately, the distribution of users was not consistent between all source accounts. The final sample included only one user from American Idol and 48 users from Rachel Maddow. At the group level, however, a chi-square test revealed no significant differences in user statistics.

<sup>5</sup> Barberá (2015) applied similar filters such that he only modeled users who posted statuses more than 100 times, posted at least one status in the previous six months, and had at least 25 followers. One notable difference is that Barberá (2015) also limited users to those who followed at least three political accounts. By expanding the range of potential elites, it is possible to examine whether users who follow entertainment accounts tend not to follow political elites.

Table 3.2: User summary statistics by source accounts

Source account	$N$	Followers	Friends	Statuses
<b>D-partisans</b>				
HuffPostPol	271	403.49	698.14	4,632.12
maddow	48	353.56	534.58	2,633.27
paulkrugman	327	389.35	689.95	3,951.07
Salon	354	400.79	716.55	3,946.75
<b>R-partisans</b>				
DRUDGE_REPORT	295	355.80	627.45	3,780.40
foxnewspolitics	212	341.88	629.04	3,890.95
SarahPalinUSA	301	376.49	675.42	3,347.28
seanhannity	192	375.44	611.85	3,936.49
<b>E-non-partisans</b>				
AMC_TV	471	331.23	713.19	4,197.57
AmericanIdol	1	291.00	1,152	742.00
SInow	211	398.43	655.28	4,399.24
survivorcbs	317	304.25	589.62	3,983.47

### 3.3.4 Data wrangling

Data were collected by recording friend networks for all users in the sample. This means a list of user Id’s—representing the followed accounts—were recorded for each user in the sample. For the sake of clarity, accounts followed *by* users will be referred to as *elites* and whether users in the sample chose to follow elites will be referred to as *follow decisions*. In order to estimate the partisanship of elites, the total number of positive follow decisions made by users sampled from each source account. Follow-decisions were aggregated for each elite (rows) and across twelve source accounts (columns), creating a frequency table to identify patterns in the accounts followed by sample users across each source account. For instance, if user A and user B were sampled from Rachel Maddow’s account, and both users also followed Barrack Obama and Hillary Clinton, while user C and user D were sampled from Sean Hannity’s account, but only C followed Obama and only D followed Hillary, then the frequency table would consist of two rows (one observation for each elite; in this case, Maddow and Hannity) and two columns (one variable for each source account; in this case, Obama and Hillary), with the respective counts for the first row (2, 2) and second row (1, 1)

falling under each of the respective columns. Accounts were considered “elites” and included in the frequency table if they (a) were followed by users sampled from at least two different source accounts, (b) had at least 400 followers, and (c) maintained unprotected, or public, Twitter accounts.<sup>6</sup> On average, the elites ( $N = 28,855$ ) had over half a million followers ( $M = 572,120.20$ ,  $SD = 2,341,635.00$ ), roughly sixteen thousand friends ( $M = 16,140.94$ ,  $SD = 66,969.80$ ), and posted over thirty-thousand statuses ( $M = 30,352.55$ ,  $SD = 72,535.85$ ).

---

<sup>6</sup> At the time of writing, user accounts on Twitter were set to unprotected/public by default.

## 3.4 Results

### 3.4.1 Overview of tests

The current study introduces a new network-based approach to measuring partisanship. To demonstrate the method, the partisan composition of user networks on Twitter were examined in three ways. First, to estimate the number of observed components in the data eigenvalue decomposition of observed follow-decisions was compared with those from a parallel analysis. Second, following a principal component analysis (PCA) with oblique rotation, the factor loadings of different source accounts across the three rotated components were examined. Finally, using regression scores from the rotated PCA solution, the validity of estimates for popular elite accounts were evaluated.

### 3.4.2 Parallel analysis

First, using the dimensions of the frequency table described above ( $N$  obs = 28,855;  $N$  vars = 12), a parallel analysis was used to examine component structure. Based on comparisons between observed and simulated eigenvalues, the parallel analysis suggested a three-component solution was best given the data. The observed and simulated eigenvalues can be seen in Figure 3.1. The parallel analysis suggests follow decisions were comprised of three components. Whether the arrangement of source accounts and components were consistent with the assumed partisan orientations, however, require further analysis of the observed components.

### 3.4.3 Rotated components

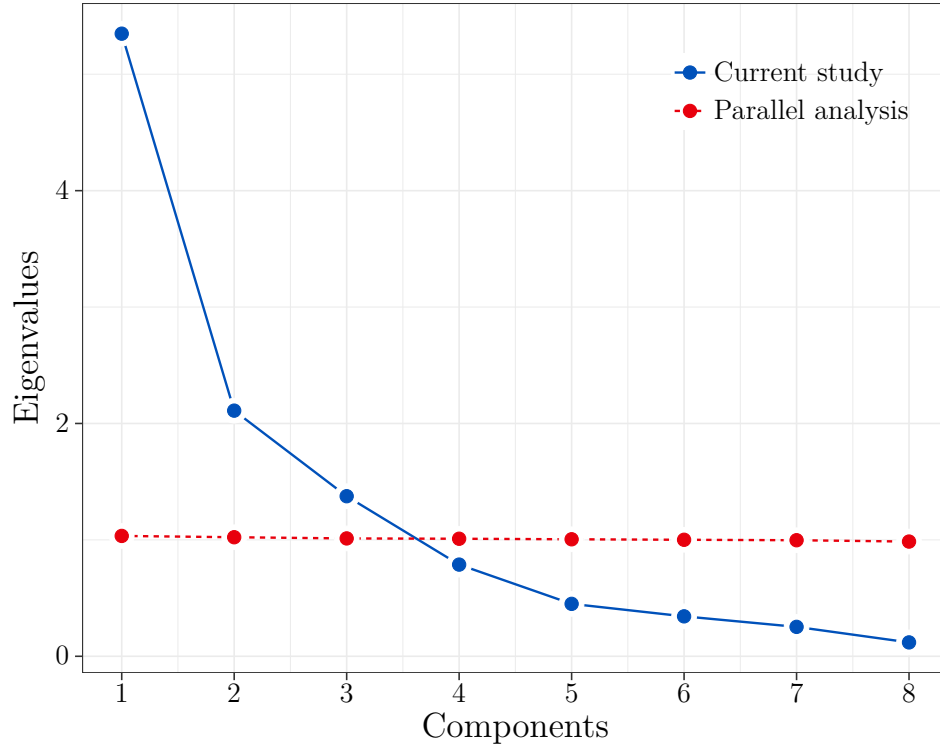
Second, a principal component analysis (PCA) was conducted followed by an oblique rotation<sup>7</sup>. Components were rotated<sup>8</sup> in much the same way as would be done in factor analysis

---

<sup>7</sup> The specific rotation method, “oblimin”, is made available in the R packages GPArotation (Bernaards & I.Jennrich, 2005) and psych (Revelle, 2016).

<sup>8</sup> Rotated components are not technically components at all; the terminology has been adopted to reflect contributions from both principal component analysis and factor analysis traditions.

Figure 3.1: Eigenvalues of follow decisions by users sampled from twelve source accounts



because, although the data do not contain measurement error, follow decisions do not occur exclusively as a function of partisanship. Meaning, it is more important to extract shared variance in follow decisions between source account groups than it is maximize the amount of total variance explained by the components. The goal is not to reproduce *all* of the variance between the groups. Rather, it is to isolate and extract the *shared* partisan dimension in follow decisions on Twitter.

Principal component analysis of the data yielded strong support for the theorized three component model,  $\chi^2(25) = 9,108.82$ ,  $p < .001$ , root mean square residual (RMSR) = 0.054. Estimates of the loadings, variances, and other fit statistics are included in Table 3.3. A spatial depiction of pattern loadings can be found in Figure 3.2 and the correlations between between can be found in Table 3.4.



Table 3.3: Standardized rotated estimates of follow decisions made by users sampled from twelve source accounts

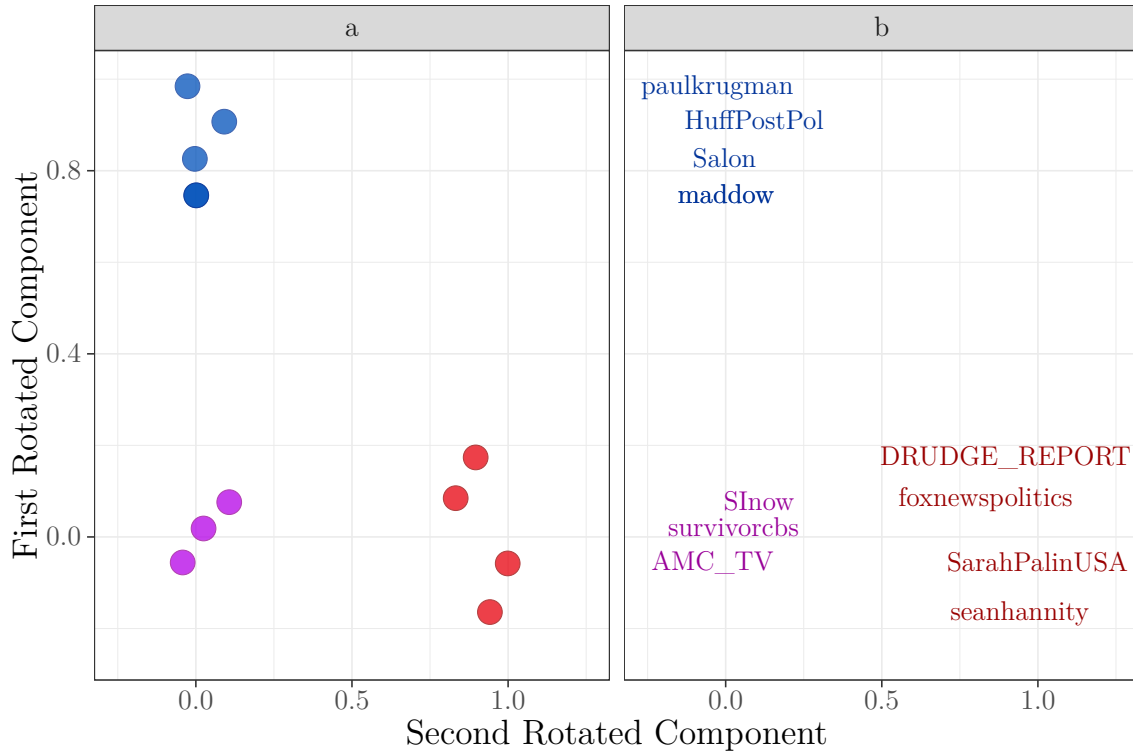
Variable	1	2	3	Comm.	Uq.
SarahPalinUSA	<b>1.00</b>	-0.08	0.02	0.95	0.05
seanhannity	<b>0.94</b>	-0.11	0.04	0.83	0.17
DRUDGE_REPORT	<b>0.90</b>	0.17	-0.03	0.94	0.06
foxnewspolitics	<b>0.83</b>	0.13	-0.02	0.79	0.21
paulkrugman	-0.03	<b>0.98</b>	-0.05	0.92	0.08
Salon	0.00	<b>0.98</b>	-0.06	0.91	0.09
HuffPostPol	0.09	<b>0.91</b>	0.05	0.93	0.07
maddow	0.00	<b>0.75</b>	0.29	0.79	0.21
survivorcbs	0.02	0.02	<b>0.84</b>	0.73	0.27
AMC_TV	-0.04	-0.01	<b>0.81</b>	0.65	0.35
SInow	0.11	0.03	<b>0.56</b>	0.37	0.63
AmericanIdol	-0.05	0.17	0.21	0.09	0.91
SS loadings	3.46	3.51	1.91		
Eigenvalues	5.38	2.13	1.39		

### 3.4.4 Applied validity

Third, the validity of accounts (identified via handle and “name” variables) were examined in relation to the strongest and weakest loadings for each component—after all, evidence of three dimensions does not necessarily mean follow decisions are shaped by partisanship. A visual inspection of the partisan orientation of several elite accounts, which is presented in Figure 3.3, makes it clear that user networks appear, at least on face, remarkably reflective of partisan orientations of well-known political elites.

For comparison, Pew Research Center’s estimates of media trust by partisan group are provided in Figure 3.4 (Mitchell et al., 2014). Although network-based measurements of partisanship differ from attitude-based metrics of ideology, the two measurements should be positively correlated. One way to validate the network based measurement, then, is to isolate several of the same sources estimated by Pew Research Center. As can be seen in Figure 3.4, network-based estimates of partisanship seem consistent with Pew’s estimates

Figure 3.2: Rotated estimates made by users sampled from twelve source accounts



derived from self-report and attitude-based measures.

One notable difference between the two approaches, however, is the degree to which New York Times and Fox News register at the extremes. In the graphic provided by Pew, the top half of Figure 3.4, New York Times appears to the left of CBS News, PBS, and Huffington Post, but to the right of Slate and the New Yorker. Similarly, Fox News appears to the right of Yahoo News, but to the left of Drudge Report, Brietbart, and The Blaze. In contrast, the network-based measure, shown in the bottom half of Figure 3.4, found that among those sources estimated by both Pew and the current study the New York Times and Fox News were by far the most extreme examples of the two competing partisan orientations.

Not only are the accounts commonly-referenced examples of well-known partisan people or organizations, but two of the most extreme scores, as depicted in Figure 3.3, happened to be the Republican and Democratic candidates for president at the time of data collection.

Table 3.4: Rotated component correlations

Variable	1	2	3
1. Republican	1.000	—	—
2. Democrat	0.378	1.000	—
3. Entertainment	0.230	0.337	1.000

A visual inspection of the degree to which certain accounts loaded on the non-partisan component compared to the degree to which the same accounts loaded on *partisan* components (democrat or republican), presented in Figure 3.5, reveals similarly striking results. The distinction between partisan and non-partisan appears, which was theorized to occur as a function of entertainment-focused media choices, adds additional evidence of the role of partisanship in the organization of user networks on Twitter.

While the rotated component scores for elite accounts reported above should satisfy any concerns about the validity of the results presented here, a more empirically-driven approach should remove any remaining doubts as to the influence of partisanship in the organization of Twitter user networks. Using the source-accounts associated with each component, follow decisions made exclusively in relation to members of the U.S. Congress were examined. Using DW-Nominate scores as an indicator of partisanship (e.g., [Hetherington, 2001](#)) allows for a more empirical approach to comparing the partisan patterns among clustered user groups. The results are, once again, striking. The mean number of follow-decisions targeted toward congressional elites (DW-Nominate metric) compared to the mean number of follow-decisions targeting Twitter elites (network-based approach) for each of the partisan sample groups are provided in Table 3.5. As provided in Figure 3.6, follow decisions for democrat (blue), republican (red), and moderate/entertainment (purple) users in the sample appear entirely consistent whether in relation to elites derived from the rotated PCA (left-hand column in Figure 3.6) or from empirically validated partisan scores associated with current members of U.S. congress.

Table 3.5: Mean number of follow decisions of congressional and Twitter elites by partisan sample group

Target of follow decisions	Democrat group	Republican group
R-Congress	0.67 (.33)	2.45 (.80)
D-Congress	1.35 (.67)	0.59 (.20)
R-Twitter	49.59 (.25)	158.68 (.71)
D-Twitter	149.58 (.75)	64.30 (.29)

Results from all three analyses—parallel analysis, principal component analysis with rotated loadings, and theoretical and empirical inspection of partisan estimates of elite accounts—make it clear that user networks on Twitter reflect partisan affiliations between users. The results suggest that partisanship not only explains differences between partisan-republican and partisan-democrat users, but between politically engaged users and disengaged users.

Figure 3.3: Rotated component scores of elite accounts

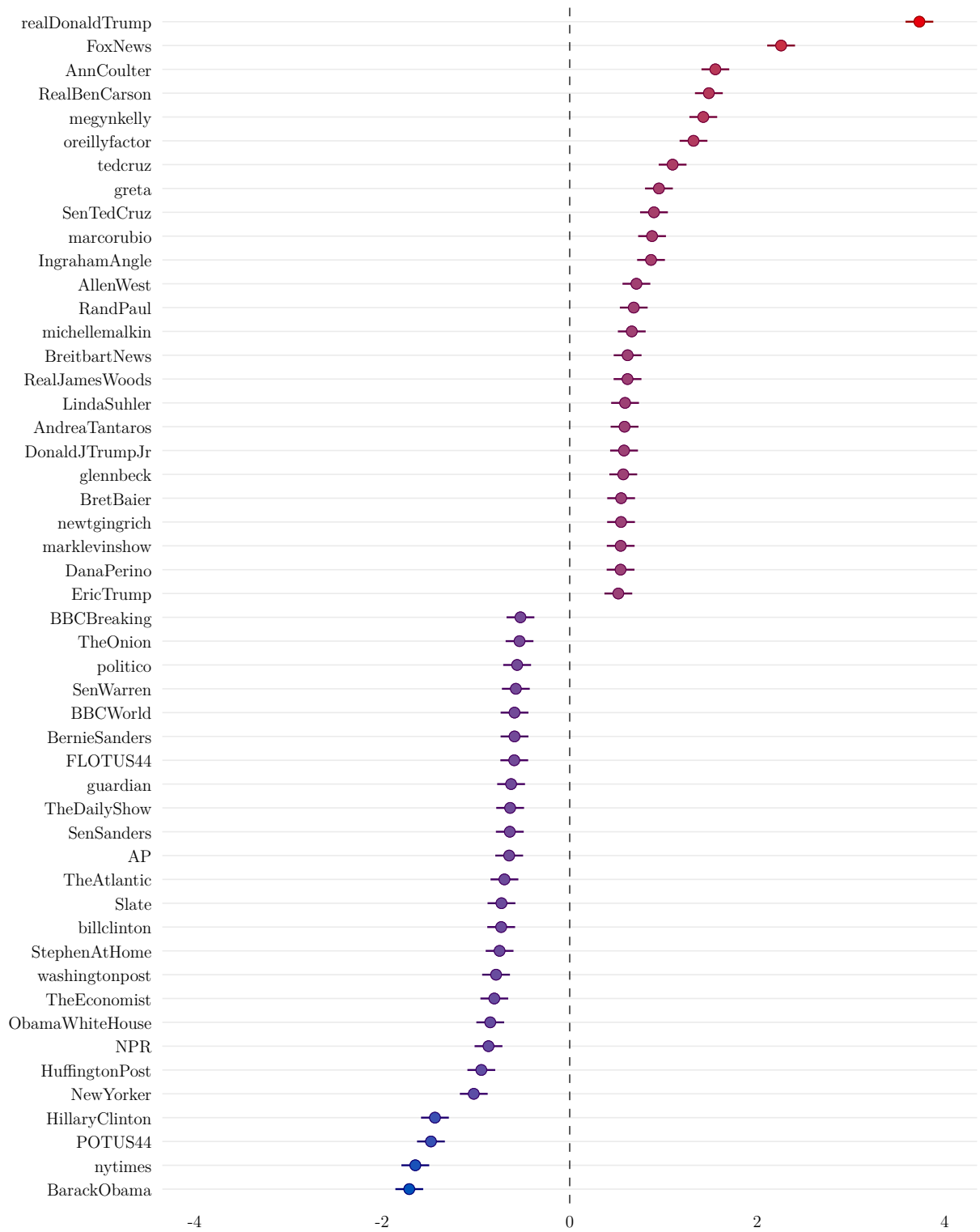
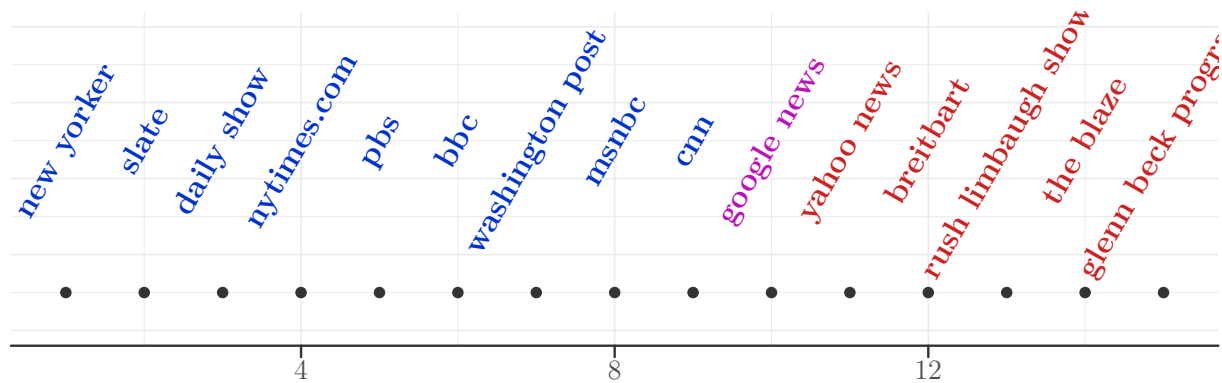


Figure 3.4: Estimate of partisan orientation of each elite's network

## Estimated partisanship of each source's audience

Partisan orientation of sources estimated by Pew



## Estimated partisanship of each source's network

Partisan orientation of elites based on follow decisions of Twitter users

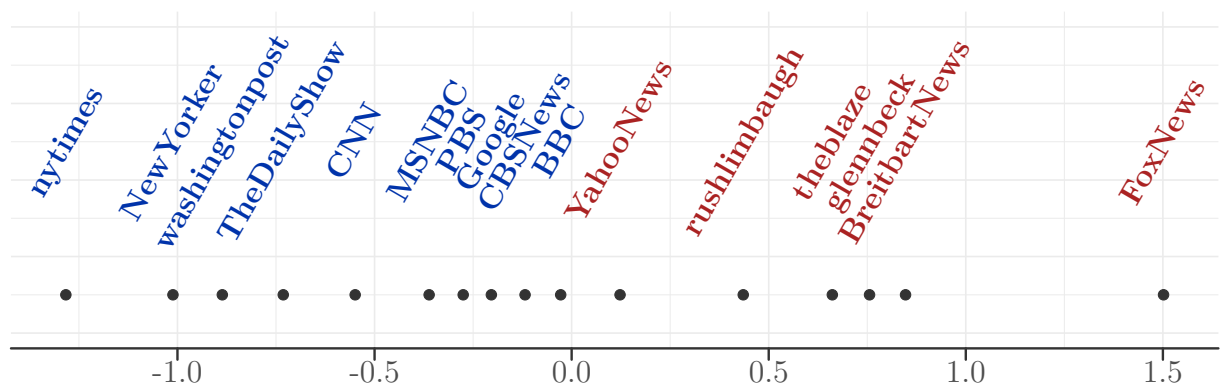


Figure 3.5: Rotated component scores of elite accounts

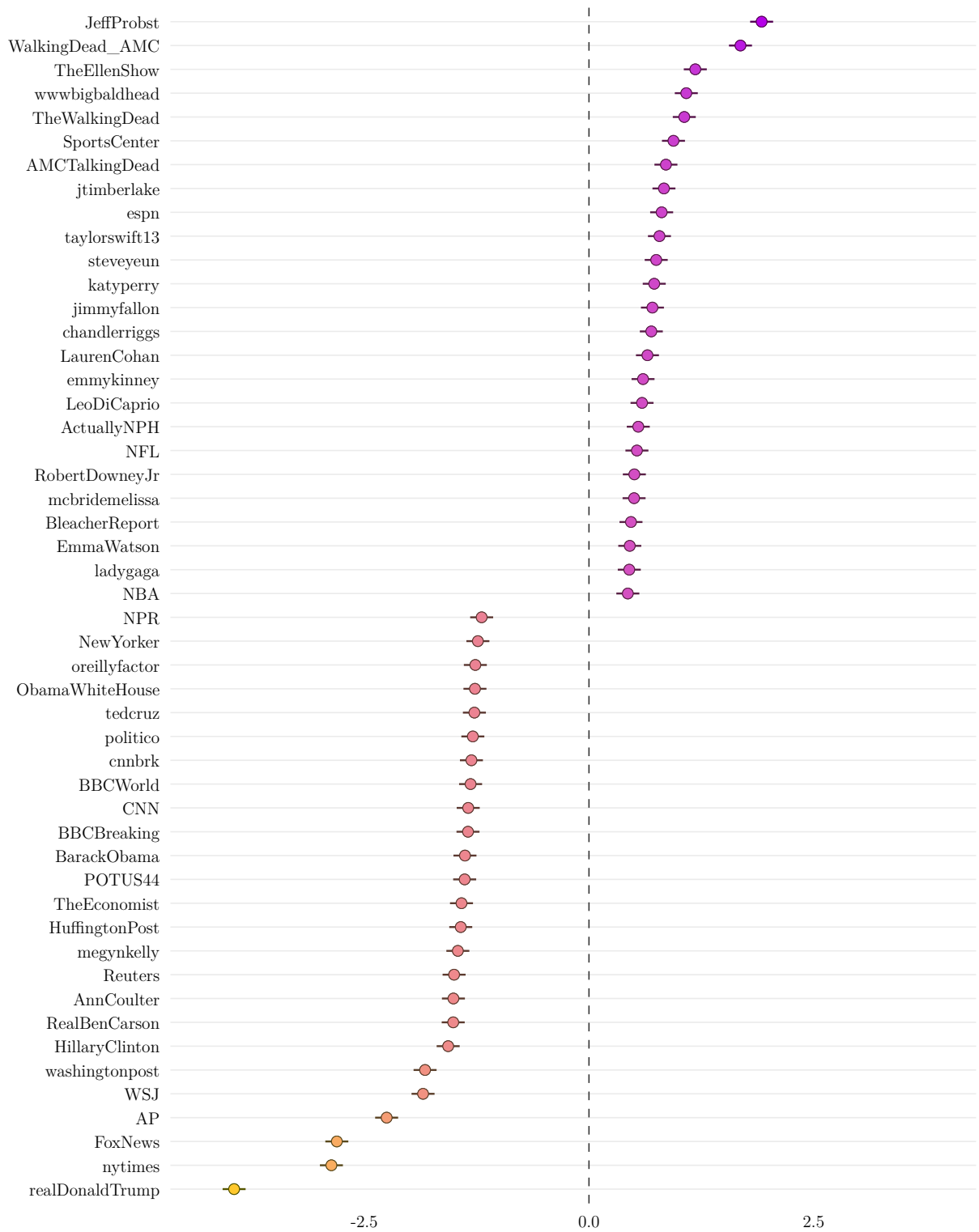
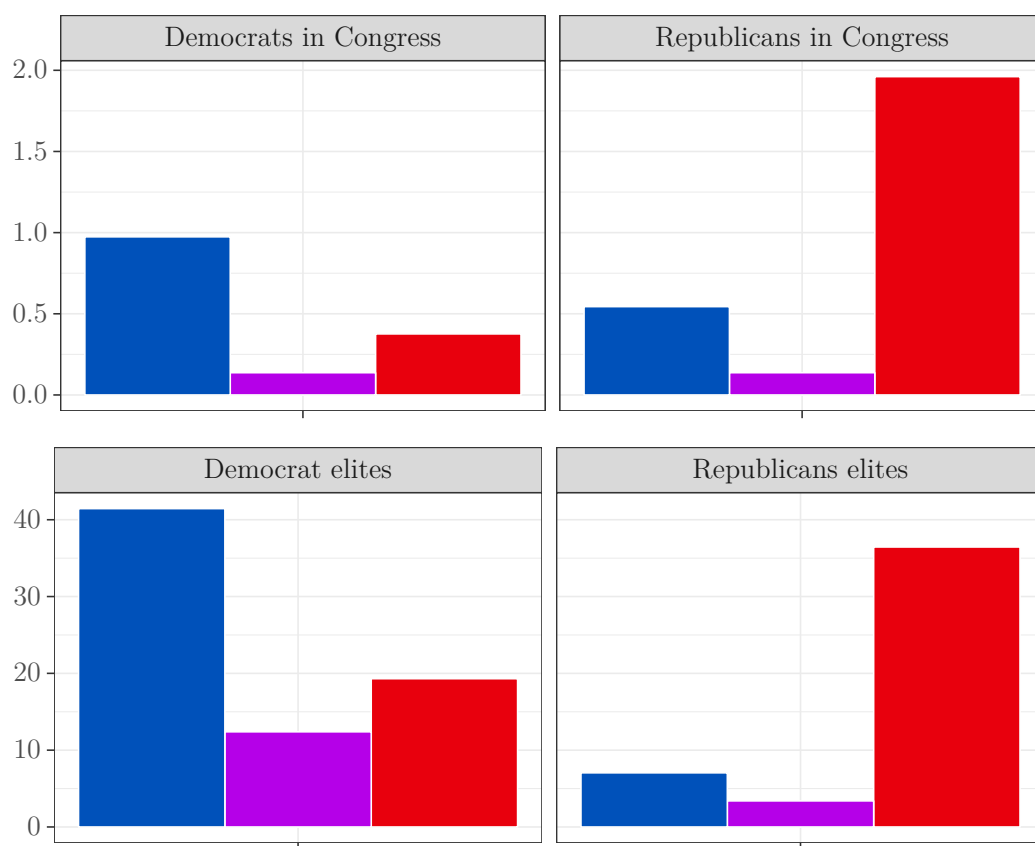


Figure 3.6: Comparing estimates of follow decisions targeting congressional elites with estimates of follow decisions targeting elites determined by a network-based approach





## 3.5 Discussion

Using follow decisions of randomly sampled users on Twitter, the current investigation sought to estimate the partisan orientation of political elites. In contrast to previous approaches, which center on user attitudes or media content as indicators of political ideology, a network-based approach to estimating partisan orientations highlights the affiliative-nature of partisanship. When leveraged with real-time data systematically accessed via Twitter’s API, the current method provides an empirically-grounded approach to estimating partisanship which is derived purely from political associations and which is free of measurement error.

To demonstrate the proposed network-based approach to estimating partisanship, three sets of analyses were conducted. First, a parallel analysis was simulated using the dimensions of the elite-level frequency data. Results from the parallel analysis suggested variance in follow decisions yielded three components. This finding was consistent with a visual inspection of eigenvalues. While these results provide support for the assumption that partisanship would explain variations in follow decisions of three groups of users, they do not go far enough to suggest these users can be distinguished by the presumed partisan orientation of the source accounts. This requires additional analysis of the observed eigenvalues, which was conducted in the second analysis.

Next, a principal component analysis (PCA) with oblique rotation was conducted. Examination of factor loadings of follow decisions for each source account revealed the source accounts clustered into partisan democrat, partisan republican, and non-partisan entertainment groups as expected. This suggests that user networks sampled from well-known partisan and non-partisan media sources can be expected to, on average, reflect distinct partisan orientations. This also means this method provides a great deal of flexibility and responsiveness. Researchers can make a few reasonable, or conservative, assumptions in a nearly infinite combination of ways about well-known partisan sources and, from that, measure the partisanship of a diverse and wide range of media sources.

Finally, using the rotated component scores generated during the PCA, consideration

was given to the applied validity of the network based approach. Evaluations of the applied validity revealed that in addition to elites being distributed by partisan orientation—that is, between democrats and republicans—these network-based estimates of partisanship correlate with attitudinal/self-report based estimates of political ideology. And with perhaps the most validating evidence of this method, results finally suggested that the pattern of follow decisions made in reference to active legislators holds for for follow decisions made in reference to entertainment and partisan elites more generally.

### **3.5.1 Measuring gradual change over time**

The network-based approach demonstrated here at a single point in time can also fill a glaring hole in the study of communication and media as it relates to the analysis of change. Many, if not most, of the constructs studied in communication and media research change over time. Change often occurs within subjects across multiple observations, which in academic research terms, translates to high costs in time, resources, or capital and high rates of attrition, or partisan dropout. We observe the obvious indicators of change over time—when a senator switches political parties, when people reach certain age-related milestones, when people get married, and when you realize MTV no longer plays music videos—but researchers rarely attempt to measure the subtleties, or different magnitudes, of change. When constructs are operationalized at the network-level, however, change can be systematically measured and analyzed in real-time among a wide range and large number of users.

### **3.5.2 Using network-based approaches in other contexts**

The findings presented here make an empirical contribution to our understanding of political partisanship on Twitter, but they also demonstrate one of many potential uses of a network-based approach to measurement. This approach could be extended to examine partisanship of any [public] population on Twitter. Applications for the network-based approach demonstrated here can extend well beyond the realm of politics. For instance, one could measure

education or socio-economic status by classifying accounts as indicators of different industries (lower-wage, manufacturing, technical, professional industries, etc.), levels of education (professors, medical doctors, etc.), or professional or technical industries as well. Assuming a defensible classification method is used, the mapping of user networks would follow much in the same way as presented in the current investigation. Other possible uses include identifying users who like certain sports teams, play a genre of video games, read certain books, speak multiple languages, identify with a country or geographical area, belong to various online communities, engage in certain hobbies, express certain movie preferences, etc. The possibilities are nearly endless. Essentially, anything a user could possibly share or otherwise reveal about themselves on Twitter could be a candidate for classification.

### 3.5.3 Limitations

Estimating partisanship of media sources via user networks on Twitter is not without limitations. Several are worth noting. First, although Twitter activity correlates with human activity in a geographical sense, Twitter users are not representative of the general population ([Garrett & Stroud, 2014](#)). They tend to be younger, more educated, and slightly more urban than the general population ([Barberá & Rivero, 2014](#)). Although perhaps less of a concern since the current study did not sample from or analyze user content, such as tweets or profile descriptions, research also suggests certain groups, such as conservatives, men, and people with strong ideological preferences, are overrepresented in political discussions on Twitter ([Barberá & Rivero, 2014](#)). Nevertheless, estimates of partisanship presented here reflect network associations among Twitter users who follow well-known, U.S.-based partisan and non-partisan accounts, and not the general population.

Second, the network-based approach, as demonstrated here, assumes follow decisions are the only mechanism through which users manage network connections. In truth, Twitter users have multiple options, beyond follow decisions, at their disposal. For instance, instead of “following” an account, users can navigate to the accounts profile page and, as

long as the account is public, get all the same information. This means users can access information from public accounts without technically “following” those accounts. However, given the current investigation’s definition of partisanship—as affiliative, or network based—distinguishing profile visits from follow decisions still seems valuable. Further, if navigating to an account home-page represents the circumvention method for following-decisions, then Twitter’s “mute” option represents the circumvention method for unfollowing-decisions. Twitter provides users with an option to mute, or hide, select accounts. The mute feature has the same effect as unfollowing insofar as the user is no longer exposed to posts sent from the targeted account. As was the case with the following-like circumvention method, however, the mute option still leaves in place the explicit network connection between users. Meaning, user behaviors that circumvent some of the assumptions made about connected users leave in tact the key characteristic on which the current study focuses.

Third, although Twitter provides thorough and accessible documentation for interacting with its APIs, collecting Twitter data using open-source software still requires at minimum either a cursory understanding of Hypertext Transfer Protocol (HTTP) requests and a moderate familiarity with the R environment or an advanced understanding of HTTP requests and a moderate familiarity with some other programming language. In addition to these technical barriers to entry, the data collection process can be tedious due to Twitter’s API rate limits. For example, gathering data on all the accounts followed by 3,000 Twitter user (the same number used in this manuscript) takes a minimum of 50 hours<sup>9</sup>.

Taken together, the analyses produced strong evidence that user networks on Twitter are organized by partisanship. Further, analyses presented here demonstrate both the value and potential in network-based approaches to estimating partisanship using Twitter data. Future research should extend this approach to other media platforms as well as international contexts. Given its growing role in political and consumer research, researchers should also explore machine-learning applications with the purpose of predicting partisanship in real-

---

<sup>9</sup> Not to mention, this also assumes an uninterrupted internet connection.

time.

# Chapter 4

## Analyzing change in network polarization

*Abstract:* The growing influence of social media in an era of media fragmentation has amplified concerns of hyper-partisanship, or political polarization. Few studies have analyzed selective exposure on Twitter using longitudinal panel data. This study examines change in network polarization on Twitter during a highly contested general election. User networks ( $N = 3,000$ ) randomly selected from followers of well-known partisan and entertainment accounts, were recorded via Twitter's API seventeen times in the 7 months leading up to the election. Results suggest that partisan users form highly partisan networks on Twitter, while moderate, or the less engaged, users continue to mostly avoid politics.

*Keywords:* change, longitudinal, networks, selective exposure, political polarization, twitter

The growing influence of social media in an era of media fragmentation has heightened concerns of hyper-partisanship, or political polarization. With the potential for algorithms to amplify the homogeneity of user networks, there is a growing concern that social media are contributing to polarization via partisan selective exposure and online echo chambers (Colleoni et al., 2014). Indeed, if new media platforms exacerbate political divisions to the point where people no longer maintain diverse channels of communication, then self-governance, public deliberation, and respect for pluralism may no longer be sustainable features of our democracy (Mancini, 2013). However, despite a common perception among scholars and pundits of a more polarized general public in recent years, evidence of political polarization is ambiguous (Prior, 2013). Upon reviewing the literature on political polarization, Prior (2013) concluded that empirical analysis has been, “severely hampered by a seemingly simple problem: we do not know how many and what kind of people are exposed to which messages” (p. 102). As such, understanding how partisan selective exposure operates on a popular social-networking platform like Twitter could tell us a lot about political polarization in the current media environment.

The purpose of the present study is to better understand change in political follow-decisions on Twitter during a high-profile, contentious general election year. Toward this end, the current study tracks a random sample of Twitter users during the 2016 general election. The investigation makes three main contributions. First, it provides an empirical account of political polarization on social media during an election using a unique and rich data set from a large, probability sample of 3,000 randomly selected Twitter users. The data collection process, which used Twitter’s API to record—without measurement error—user networks at seventeen different time points throughout the 2016 general election, offers the first glimpse of change in actual polarization behaviors over the course of an active election. Second, the study sheds light on previously under-studied dynamics of user networks *over time* on Twitter. And, third, it provides powerful evidence of why we need to take seriously the influence of social media on civic engagement and the political polarization process.

Overall, the study makes a valuable contribution to our understanding of change in user networks and political polarization during an acutely polarizing-time in electoral politics (Gramlich, 2016).

## 4.1 Measuring polarization

The consensus in the literature suggesting political polarization is the distance between competing political orientations quickly breaks down when scholars start defining “political orientation” or, especially, “distance.” By evoking a metaphor of “polarization,” consideration of movement, i.e., distance, away from the center and toward the poles naturally follows. What this means in political contexts, however, is not simple. For the most part, concerns of polarization in the literature focus on the dangers of not communicating with people from the other side (e.g., Arceneaux et al., 2012; Fischer et al., 2005; Garrett, 2009; Kim, 2011; Stroud, 2010). Indeed, concerns undoubtedly reflect contemporary democratic principles of free expression, debate, and public deliberation, but none of these principles readily translate into physical distance. Rather, most definitions of polarization in political communication research fall into one of five different categories—voting records (e.g., Bartels, 2000; Hetherington, 2001), issue positions and ideologies, (e.g., Abramowitz & Saunders, 1998; Baldassarri & Gelman, 2008; Lee et al., 2014), attitudes (e.g., Lau et al., 2016; Lee et al., 2014; Westfall et al., 2015), media exposure (e.g., Arceneaux et al., 2012, 2013; Levendusky, 2009), and social networks (e.g., Himelboim et al., 2013b; Kim, 2011). Although each category has its own strengths and weaknesses, which will be discussed in more detail in the following section, across all categories scholars have been limited in their ability to exploit the obvious parallels between polarization and social networks in particular. In other words, although partisanship is formed and maintained via the organization of political social networks, technological limitations and convenience of self report has produced definitions of partisanship that rarely examine *networks* for evidence of political polarization. Through



the use of recent advancements in technology, this study seeks to address this gap in the research.

#### 4.1.1 Voting records

The clearest evidence of polarization can be found among political elites, specifically elected officials. For the past several decades, considerable evidence suggests political candidates and elected officials have become more polarized (Abramowitz & Saunders, 1998; Hetherington, 2001; Layman & Carsey, 2002; Levendusky, 2009; McCarty et al., 2009). One would assume such obvious trend of elite-level polarization would contribute to *mass polarization*, or the political polarization of the general public. While several studies have supported this assumption (e.g., Abramowitz & Saunders, 1998; Hetherington, 2001) other studies found evidence of *party sorting*, the proportion of people affiliated with one of the two major parties, but not mass polarization (Fiorina & Abrams, 2008; Hill & Tausanovitch, 2015). But it is also possible that neither explanation is correct. In other words, the polarization of legislators, for example, could be due to gerrymandering (Carson et al., 2007; Mann, 2007; McCarty et al., 2009) or even evolutions in political strategy (Ansolabehere et al., 2010; Theriault & Rohde, 2011). Ultimately, voting records of legislators alone do not translate to polarization of the public. Research must instead continue to look at behaviors and communication across the public to understand mass polarization.

#### 4.1.2 Issue positions and attitudes

A second, predominant category of research on political polarization places an emphasis on issue positions or political ideology (e.g., Abramowitz & Saunders, 1998; Baldassarri & Gelman, 2008; Lee et al., 2014). Therefore, across the research issue-position polarization and ideological polarization are similar in that they both attempt to describe the degree to which people hold competing political views. However, there are problems with these definitions of polarization. For instance, they assume political sentiment reflects some objective reality

of political opinions and ignore antagonistic expressions of of political competition. Even if it were possible to accurately measure the degree to which political opinions differ, there is little reason to believe perceived differences would correspond with objective differences. In other words, political opinions and ideologies do not adhere strictly to some certain set of rules or regulations. Research on political cognitions makes it clear the role of partisanship should not be understated (Nyhan & Reifler, 2010; Westen, 2008). No doubt insights can be gleaned from studying affective competition between parties (e.g., Lau et al., 2016; Lee et al., 2014; Westfall et al., 2015), but affect, or emotion, does not directly represent the “dangerous” forms of democracy associated with polarization. Some theorists even claim that having passion and heated debates are cornerstones of democratic deliberation (e.g., Mouffe, 2005).

Further, a significant challenge in the literature finds that when considered in the context of new digital media environments, operationalizing the third category of definitions—partisan attitudes—nearly always suffers from a reliance on self-report data (Prior, 2013). Difficulties in the measurement of media exposure have, consequently, highlighted the importance of accuracy and reliability in academic research (e.g., Prior, 2009a,b). And these concerns are magnified for studies that attempt to analyze change partisan media behaviors over time. The current study remedies this gap in the literature by measuring, directly, polarization behaviors of users during a highly-salient time in U.S. politics.

### **4.1.3 Media exposure**

The fourth category of research on political polarization concerns selective exposure and media exposure in an era of unprecedented media choice (e.g., Arceneaux et al., 2013, 2012; Bennett & Iyengar, 2008; Holbert et al., 2010; Prior, 2005, 2013; Stroud, 2008, 2010). Advances in digital media have lowered the costs of entry for providers of media content. As a result of lowered entry costs, highly-specialized and audience-specific news outlets have quickly populated cable networks and online domains. In the political context, the fragmen-

tation of media has lead to a considerable increase in partisan media sources ([Levendusky, 2013](#)). Given this, it is hardly surprising the political consequences of increased media choice brought on by the digital age have been linked by many to perceived increases in political polarization, especially given the increase in mass polarization seems to coincide with the expansion of media choice and the subsequent rise of partisan media (e.g., [Hollander, 2008](#); [Jones, 2002](#)). However, due to limitations in current approaches to the measurement of media exposure, evidence of political polarization via partisan selective exposure in new media environments remains unclear ([Prior, 2013](#)).

## 4.2 Social networks

### 4.2.1 Social media

In today's new media environment, understanding the role of media choice is essential to understanding media effects. It is therefore important to examine how emerging media technologies, in particular, shape the flow of information. So far, the most important of these emerging technologies is social media. Use of social media platforms, such as Facebook, Twitter, Snapchat, and Instagram, has become so widespread that American adults are arguably more likely to use social media—[Pew Research Center \(2017\)](#) estimates 76% of adults use social media—than read a book—[YouGov \(2013\)](#) estimates 72% of all adults read a book in the previous year. Over the past several years, social media has become a fixture of the digital media information environment. Even traditional media outlets are now adapting to a new media environment ([Gleason, 2010](#)). Pew research estimates that over 60% of American adults get news from social media ([Gottfried & Shearer, 2017](#)), and political communication research has confirmed the importance of social media in political contexts as well. For instance, researchers have linked social media use to voting ([Bond et al., 2012](#)), political expression ([Warner et al., 2012](#)), civic engagement ([Gil de Zúñiga et al., 2012](#)), and even political revolutions ([Tufekci & Wilson, 2012](#)).

### 4.2.2 Partisan selective exposure

In the current political landscape, it is clear that politics is largely dictated by political parties (Gramlich, 2016). Scholars generally agree that partisan sorting, or increases in partisan consistency, has increased in recent years (e.g., Hetherington, 2001). However, research is needed to examine how competition between political parties spills over and manifests among the public. For example, we know that partisanship can trump political ideology (Glaeser & Sunstein, 2013; Kahan et al., 2013; Kahneman, 2011; Warner & McKinney, 2013), but we do not know how it effects exposure to diverse viewpoints. Because partisanship is about political affiliations, and not about ideologies or opinions, polarization should therefore be understood at the network-level (e.g., Himelboim et al., 2013b; Kim, 2011)..

Many selective exposure behaviors on social media such as *follow-decisions*, or choosing to follow another account, are influenced by factors other than political ideology—e.g., interpersonal relationships or self-presentation goals. Analyzing survey data collected by Pew Research, Bode (2016) found politically motivated *unfriending* occurred among fewer than 10% of respondents. However, it was most common among people who talked more frequently about politics, held strong ideological beliefs, and encountered more information on social media. The Pew survey did not ask about Twitter, specifically, and most of the items borrowed terminology from Facebook and framed the questions to be about the respondents’ “friends” on social media. Interpersonal dynamics are especially salient on platforms like Facebook where user networks largely reflect interpersonal relationships that extend offline (Wilson et al., 2012) and where follow-decisions must be reciprocal (one user initiates a friend request, and the other user chooses to accept or deny it). Perhaps the biggest problem with the Pew survey data is that it relied solely on users to self report both their frequency and their motivations for unfriending. Media exposure research makes it clear that participant recall is not entirely reliable. It is also likely not socially desirable to “unfriend” someone or to admit to appear to be “thin skinned.” With this in mind, the current study examines a social media platform that should be more vulnerable to partisan

motivations as explained in the following section.

Although the current study does not capture user attitudes, there are nevertheless reasons to suspect partisanship will systematically influence decisions of Twitter users. Building on the assumption that all reasoning is motivated (Kunda, 1990), the model of motivated skepticism, for example, explains selective exposure decisions in the context of political information (Lodge & Taber, 2000; Taber & Lodge, 2006). This theory of motivated reasoning suggests media exposure decisions are driven by two competing goals—accuracy and partisanship. Accuracy goals reflect the desire to seek out valid, or correct, information. Partisan goals reflect the desire to defend one’s [prior] beliefs. Thus, motivated reasoning suggests that when people encounter new data, they process it along with their prior attitudes. This means people are always updating, and not purely evaluating, information. The degree to which new information influences beliefs relative to prior information depends on the strength and motivation of those priors. Given what we already know about political partisanship, new political information rarely overwhelms prior partisan commitments.

### 4.2.3 Twitter

When comparing the various social media platforms, there is reason to believe that for users on Twitter, whether they decide to follow political accounts should be relatively unburdened by interpersonal considerations (Colleoni et al., 2014). Unlike Facebook, follow-decisions on Twitter are uni-directional. Users can decide to follow any public account (Twitter allows users to opt-in to *protected* accounts, which means their time-lines are not publicly available and followers must be granted permission by the original user) whether or not the public account follows in return. A quick examination of political Twitter accounts reveals the vast majority of political elites do not *follow back*. This suggests most users do not personally know the elites they choose follow. Further, when Twitter users decide to follow political elites, communication tends to be one-directional (from elite to user). For these reasons, it seems unlikely that follow-decisions regarding well-known, or elite, accounts are driven by

interpersonal factors.

Although users often encounter news on multiple different platforms, research focusing on Twitter, in particular, offers three advantages. First, while more people report using Facebook overall, a higher percentage of people *seek out* news on Twitter (Gottfried & Shearer, 2017). And, Twitter is hardly struggling to attract users. As of 2016, roughly 16% of American adults report using Twitter (Gottfried & Shearer, 2017), which translates to roughly 40 million people in the United States alone. Given the number of users and the salience of trending information, there is reason to believe user behaviors on Twitter will vary as a function of proximity to an election (e.g., Jang & Pasek, 2015).

The second advantage to narrowing the focus of study to Twitter is the unique nature of user connections. Unlike most social media platforms, user connections on Twitter operate asymmetrically. That is, user A can follow user B even if user B chooses not to follow [back] user A. To be consistent with Twitter documentation, the current investigation refers to users who follow an account as *followers* and the users followed by an account as *friends*. So, in the example of users A and B, user A would be considered a follower of user B, while user B would be considered a friend of user A. The asymmetrical nature of follower/friend user networks means users can easily make connections with people, or organizations, they do not personally know, which makes Twitter a natural destination for affiliative expressions (e.g., Hong, 2012). Accordingly, it is not uncommon for Twitter users to follow one or more news organizations or political figures. Even the initial account creation process on Twitter, where users are asked about their interests in order to generate a list of recommended accounts to follow, encourages users to make asymmetrical connections with particularly well-known and popular Twitter accounts.

The third advantage to narrowing the focus of study to Twitter is the availability of real-time data generated by millions of users. Other social media platforms offer similar data-sharing services, but few can match the amount of data and the accompanied documentation provided by Twitter. Leveraging such large amounts of Twitter data during especially salient

times for politics, like general elections, makes it possible to examine concepts previously unaddressed in the literature—e.g., how do political networks change over time? Do politically disinterested people increasingly form new, political connections or do they “tune out” of politics entirely over the course of an election?

In light of the growing influence of social media, the information-seeking nature of asymmetrical certain user networks, and the availability of Twitter data, the current study examines selective exposure in follow-decisions on Twitter leading up to the 2016 general election.

#### **4.2.4 Previous research**

Several studies have examined political information on social media ([Barberá, 2015](#); [Barberá et al., 2015](#); [Bode, 2016](#); [Himmelboim et al., 2013a,b](#); [Larsson & Moe, 2012](#)), but more research is needed to better understand the role of partisanship in selective exposure on social media. Previous research suggests political orientations shape how people select traditional media ([Stroud, 2008, 2010](#)), but relatively little research has examined political partisanship and selective exposure on social media. Given what we know about partisanship and selective exposure in traditional media, one could assume that, on the whole, the partisan orientation of political elites should match the partisan orientation of social media users who decide to follow them. For the most part, this has been confirmed by research ([Barberá, 2015](#); [Barberá et al., 2015](#); [Himmelboim et al., 2013a,b](#)). Additionally, previous research has also found evidence of polarization among users in retweet networks during the 2010 midterm election ([Conover et al., 2011](#)), the circulation of rumors during the 2012 general election ([Shin et al., 2016](#)), and the partisan discussion networks following the 2012 State of the Union ([Himmelboim et al., 2013b](#)).

There are numerous examples of recent studies that demonstrate the tremendous research potential of social media. There are also plenty of examples illustrating the difficulties associated with content-based definitions of political partisanship and political polarization.

Using data collected via Twitter’s firehose API, [Shin et al. \(2016\)](#) analyzed over three-hundred million Tweets over the course of fifteen months in order to study how rumors circulated during the 2012 election. The researchers identified a total of 82 rumors that were mentioned on at least one major fact-checking website, with 57 of those rumors making it to the final analysis. In addition to exorbitant financial costs involved with gaining access to this kind of data (Twitter’s firehose API is not public), it took the researchers eleven months to human code over one-hundred and eighty thousand Tweets!

## 4.3 Change in polarization

### 4.3.1 Analysis of change

Research has provided convincing snapshots of political polarization on Twitter ([Barberá, 2015](#); [Barberá et al., 2015](#); [Himmelboim et al., 2013b](#); [Larsson & Moe, 2012](#)), but few studies have analyzed partisan selective exposure on network polarization on Twitter using longitudinal panel data. Of the few studies that have, even fewer addressed the concept of change. For instance, some of the most in-depth analysis on the subject of political polarization on Twitter to date involves repeated collection of data over the course of several days during an election [Barberá \(2015\)](#); [Barberá et al. \(2015\)](#), but refrains from examining cross-sectional research questions. To date, no study has used repeated measures of follow-decisions made by ordinary users on Twitter to model between-user *and* within-user change during a major election. Without an understanding of how these behaviors evolve over time and especially during elections, it is impossible to effectively assess the political implications of social media use.

### 4.3.2 Political salience and the election

In its simplest form, the hypothesis forwarded here suggests that change in network polarization, or the number of homogeneous, or partisan-consistent, follow-decisions compared to



the number of heterogeneous follow-decisions made by a user, should occur as a function of proximity to the election and whether or not the users were sampled from a partisan rather than entertainment-centered account. Studying the trajectories of selective-exposure decisions on a popular social media platform like Twitter should shed light on the role of change in the polarization of user networks in new media environments.

Within users, one would expect to find that network polarization increases as the election approaches. Proximity to the election should also raise the salience of political orientations, driving users to increasingly add more homogeneous follow-decisions to their user networks. Not only should politics become more salient as coverage of the election intensifies, but the value placed on homogeneous political connections should increase as well. This would explain why, using ANES data collected during the 2004 election, [Stroud \(2008\)](#) found that people increasingly selected attitude consistent cable news programs leading up to the election. So, if the salience of partisanship positively correlates with proximity to the election, and if political stakes grow as the election approaches, then the effect of partisan selective exposure on follow-decisions should increase within-users leading up to the election as well. Therefore, this study tests the following hypothesis:

H1: Partisan network homogeneity, or network polarization, will increase with proximity to the election.

## 4.4 Method

### 4.4.1 Data collection

To test the theorized relationship between partisan-preferences of users and proximity to the election, a sample ( $N = 3,000$ ) of partisan and non-partisan users were randomly sampled and tracked at frequent time intervals ( $N \text{ waves} = 17$ ) in the months leading up to the 2016 election. Data collection started on June 13, 2016, shortly after the second of the two major party candidates for U.S. president—Hillary Clinton of the Democrats and Donald Trump of the Republicans—begin their party’s presumptive nominees, and ran until November 11, 2016, three days after the general election.

Data were collected using the R package *rtweet* (Kearney, 2016) to access Twitter’s public REST Application Program Interfaces (API). The population consisted of followers of several well-known partisan and non-partisan source accounts. The republican group consisted of users who followed Drudge Report (@DRUDGE\_REPORT), Fox News Politics (@foxnewspolitics), Sarah Palin (@SarahPalinUSA), or Sean Hannity (@seanhannity). The democratic group consisted of users who followed Huffington Post Politics (@HuffPost-Pol), Rachel Maddow (@maddow), Paul Krugman (@paulkrugman), or Salon.com (@Salon). Lastly, the entertainment group consisted of users who followed AMC TV (@AMC\_TV), American Idol (@AmericanIdol), Sports Illustrated (@SInow), or CBS’s Survivor (@survivor-cbs).<sup>1</sup>

A large sample of users ( $N = 60,000$ ) were sampled from the pools of followers ( $n \text{ group} = 20,000$ ) of the accounts listed in the previous paragraph. Users data was then collected on all users and filters were then applied to remove inactive users and bot accounts. Users were removed if they had fewer than 50 or greater than 1,500 followers (accounts who follow them) or friends (accounts they follow), posted fewer than 200 statuses in total, or failed to post a status on average every ten days. A final sample of users were then sampled from

---

<sup>1</sup> For more detailed explanation of the sampling, selection, and filtering methods, see Kearney (2017).

each of the filtered pools of users. The final sample ( $N = 3,000$ ) consisted of users assigned to three equally-sized groups ( $n = 1,000$ ). For a summary of user statistics for each of the three groups, see Table 4.1.

#### 4.4.2 Sample users

Followers were identified collected for all followers of each of the selected source accounts using the *rtweet* function `get_followers`. Once all followers were identified, user Id's were pooled into the three groups—the presumed groups associated with each source account—and a number of users, which was more conducive to API rate limits, was randomly selected from each group ( $N = 60,000$ ; 20,000 users per group). To filter out spam, bot, and inactive accounts, user-level [meta] data was retrieved using *rtweet*'s `lookup_users()` function. Guided by first-hand experience and previous research (e.g., Barberá, 2015; Haustein et al., 2016; Yardi et al., 2009), several filters were selected and then applied to user-level data. Users were removed if they were *protected* (set their accounts to private), had fewer than 50 or greater than 1,500 friends or followers, fewer than 200 tweets, or failed to post a status an average of once every ten days.<sup>2</sup> Finally, to ensure consistency and to make navigating Twitter's API rate limits more reasonable, an equal number of users ( $n = 1,000$ ) were sampled for each group, yielding a final sample size of 3,000 users ( $N = 3,000$ ). User summary statistics for each source account can be seen in Table 3.2. Unfortunately, the distribution of users was not consistent between all source accounts. The final sample included only one user from American Idol and 48 users from Rachel Maddow. At the group level, however, a chi-square test revealed no significant differences in user statistics.

---

<sup>2</sup> Barberá (2015) applied similar filters such that users were only modeled if they posted more than 100 statuses, sent at least one status in the previous six months, and had at least 25 followers. One notable difference is that Barberá (2015) also limited users to those who followed at least three political accounts. By expanding the range of potential elites, it is possible to examine whether users who follow entertainment accounts tend not to follow political elites.

Table 4.1: Summary statistics of sample and group

Group	Waves	N	Followers	Friends	Statuses
Dem.	16.52	971.88	396.12	736.38	4,029.10
Ent.	15.84	931.88	335.88	694.26	4,108.36
GOP	16.24	955.42	364.19	705.42	3,690.30
Mean	16.20	2,859.18	365.81	712.30	3,941.72
SD	2.84	67.93	268.84	423.89	5,332.66

#### 4.4.3 Follow decisions

The dependent variable of interest is network polarization, the degree to which the partisanship of a user’s network is homogeneous. Network polarization was measured by collapsing all observations across time and calculating a weighted estimate of partisanship for each follow-decision made by sample users in each group. To classify the partisanship of follow decisions, target accounts, or *elites*, were identified as republican, democrat, or non-partisan (entertainment). Given the sampling method used—random samples of users taken from multiple source accounts selected to represent two partisan and one non-partisan group—whether the target [account] of a follow-decision was republican or democrat was approached as an empirical question. Thus, targets of follow decisions were considered elites and included in the analysis if they were followed by at least one sample user from two different source accounts—regardless of whether the source accounts came from different partisan groups. For example, an account was included as an elite if it was followed by a user sampled from Sean Hannity’s followers and a user sampled from Sarah Palin’s followers.

A principal component analysis (PCA) with oblique rotation validated group assignments for the twelve source accounts. That is, an examination of observed clustering of follow decisions by users sampled from each of the source accounts confirmed the assumed group associations described above. Regression scores for each of the three components—democrat sources, republican sources, and entertainment sources—generated by the PCA were then used as weighted estimates of the elite accounts. At each time point, the sum of

weighted estimates for each component (partisan-republican, partisan-democrat, and non-partisan/entertainment) was calculated based on the follow decisions made by each user. Network polarization was then calculated at each time point by taking the absolute value of the difference between the sum of partisan-republican follow-decisions and the sum of partisan-democrat follow-decisions.

## 4.5 Results

By randomly sampling from well-known partisan-republican, partisan-democrat, and non-partisan entertainment accounts, and by collecting repeated measures ( $n_{waves} = 17$ ) of each user’s follow decisions in the months leading up to the election, the current study has compiled a unique data set that makes it possible to not only examine partisanship of networks *between* users but also to examine change in networks *within* users.

It was hypothesized that follow decisions during the 2016 election would vary as a function of user partisanship and proximity to the election. To account for other user-related differences, models also included several time-invariant covariates, including years since joining Twitter (account age), the number—in thousands—of tweets posted by a user (statuses), the number of accounts—in hundreds—a user follows (friends), as well as the number—in hundreds—of accounts that follow a user (followers).

### 4.5.1 Multilevel modeling

Models were estimated using multilevel modeling, which provides a framework for making inferences for both constant (fixed) and varying (random) effects (Gelman et al., 2005). The decision to use multilevel modeling was supported by empirical and theoretical reasons. As seen in Figure 4.1, there was considerable observed variance in follow decisions between users in both their starting points (intercepts) and their change over time (slopes), intraclass correlation coefficient (ICC) = 0.75—95% confidence intervals = 0.74 - 0.76. Second, it was theorized that the effect of proximity to the election on follow decisions would depend on whether users were partisan or non-partisan. Thus, a random intercept and random slope was included in all models. Results from ordinary least squares (OLS) regression and generalized negative binomial models can be found in Appendix 6.1 and Appendix 6.1.3.<sup>3</sup>

---

<sup>3</sup> Note: OLS and negative binomial model results appear largely consistent with the multilevel estimates; however, the models are not equivalent. The OLS and negative binomial models assume that changes in partisan homogeneity occur uniformly across users. Only the multilevel models allow users to vary both in starting values (varying intercepts) and in rate of change over time (varying slopes).

	Model 1	Model 2	Model 3
<b>Fixed effects</b>			
(Intercept)	18.96*** (3.42)	−12.70*** (3.45)	−16.06*** (3.47)
Account age	−1.62** (.55)	−1.90*** (.51)	−1.90*** (.51)
Statuses	−.49* (.23)	−.22 (.22)	−.22 (.22)
Followers	.00 (.46)	−.84* (.43)	−.83 (.43)
Friends	7.97*** (.06)	7.98*** (.06)	7.98*** (.06)
Weeks	.21*** (.03)	.21*** (.03)	−.09* (.04)
Partisan		52.62*** (2.29)	57.69*** (2.37)
Partisan*Weeks			.45*** (.05)
<b>Random effects</b>			
Var: User (Intercept)	4,438.60	3,709.90	3,704.21
Var: User weeks	1.84	1.84	1.79
Cov: User (Intercept) weeks	27.98	22.71	22.21
Var: Residual	38.98	38.98	38.98
AIC	251,288.99	250,816.11	250,748.21
BIC	251,373.30	250,908.84	250,849.37
Log Likelihood	−125,634.50	−125,397.05	−125,362.10
Num. obs.	33,878	33,878	33,878
Num. groups: User	3,000	3,000	3,000

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

Table 4.2: Varying intercept/slope estimates predicting change in homogeneous follow decisions

### 4.5.2 Models

Coefficients for the three estimated models predicting the weighted number of homogeneous (partisan-matching) follow decisions are included in Table 4.2. Model 1 included several level-2 predictors, including account age, statuses, followers, friends, and the number of weeks from the start of data collection. Model 2 added the partisan grouping variable—that is, whether a user belonged to a partisan group (democrat or republican) or the non-partisan group. Model 3 added the interaction of the number of weeks from the start of data collection with the partisan grouping variable (Partisan \* Weeks).

Models were compared using the change in chi-square ( $\Delta\chi^2$ ) test. The test revealed Model 2 fit the data significantly better than Model 1  $\Delta\chi^2(1) = 474.88, p < .001$ , which suggests that partisan user groups contributed significantly to explaining change in partisan follow-decisions. To determine if the follow decisions of partisan-democrat/republican users differed from those of non-partisan/entertainment users as a function of proximity to the election, Model 2 was then compared with the final model. The chi-square test revealed that Model 3 fit the data significantly better than Model 2,  $\Delta\chi^2(1) = 69.90, p < .01$ . This finding suggests the relationship between proximity to the election and homogeneous follow-decisions was conditional on the partisanship of the user, which is consistent the study’s hypothesis.

### 4.5.3 Interpreting the interaction

A visual depiction of the interaction is provided in Figure 4.2. The second facet (on the right) in Figure 4.2 shows the average number of weighted follow decisions for partisan users over time with, with the red line representing the average of the republican group and the blue line representing the average of the democrat group. As expected, the number of weighted homogeneous follow decisions of partisan users increased as the election got closer. In contrast to partisan users, the first column (left side) of Figure 4.2 shows that follow-decisions of non-partisan/entertainment users increased only slightly as the election approached. Taken together, these results support the hypothesis that homogeneous follow



decisions would vary as a function of user partisanship and proximity to the election.

Regardless of the model, when examined at the group level, there is no missing the clear preference of politically homogeneous follow decisions among Twitter users. This is perhaps best illustrated in Figure 4.3, which depicts change in partisan follow decisions—relative to the baseline partisanship of user networks at time 1—in each of the groups of sampled users during the 6-7 months leading up to the 2016 election. As expected during a highly polarizing election contest, partisan-democrat, partisan-republican, and non-partisan/entertainment follow decisions made by users in each group, which are represented in the three columns in Figure 4.4 and Figure 4.3, appeared to increase as the election got closer. The rate of partisan follow decisions, represented by the red (republican) and blue (democrat) lines in Figure 4.3, appear to vary according to the partisanship of the user group. In other words, homogeneous follow decisions—when users from a partisan group follow elites that scored highest in the *matching* partisan component—were made more often than heterogeneous follow decisions—when users from a partisan group follow elites that scored highest in the *competing* partisan component. In contrast to the variations observed in partisan follow decisions, the rate of non-partisan follow decisions, represented with the purple line in Figure 4.4, increased only slightly over time for all groups. This suggests that while all users tend to follow more elite accounts over time, the rate at which partisan users make partisan-matching follow decisions demonstrates a clear preference for partisan homogeneity in user networks.

Figure 4.1: Individual trajectories

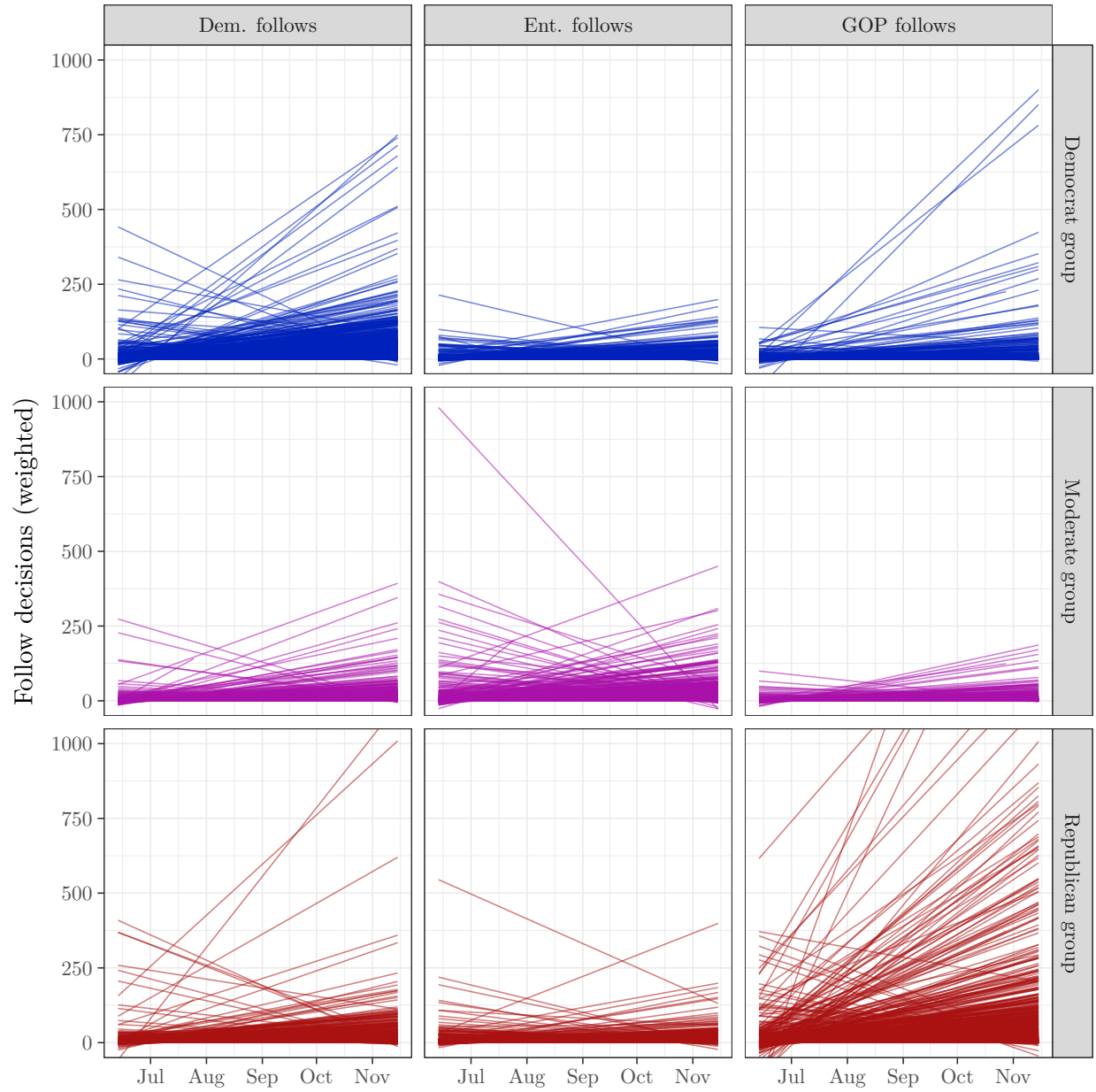


Figure 4.2: Change in number of weighted partisan follow-decisions by partisan versus non-partisan group

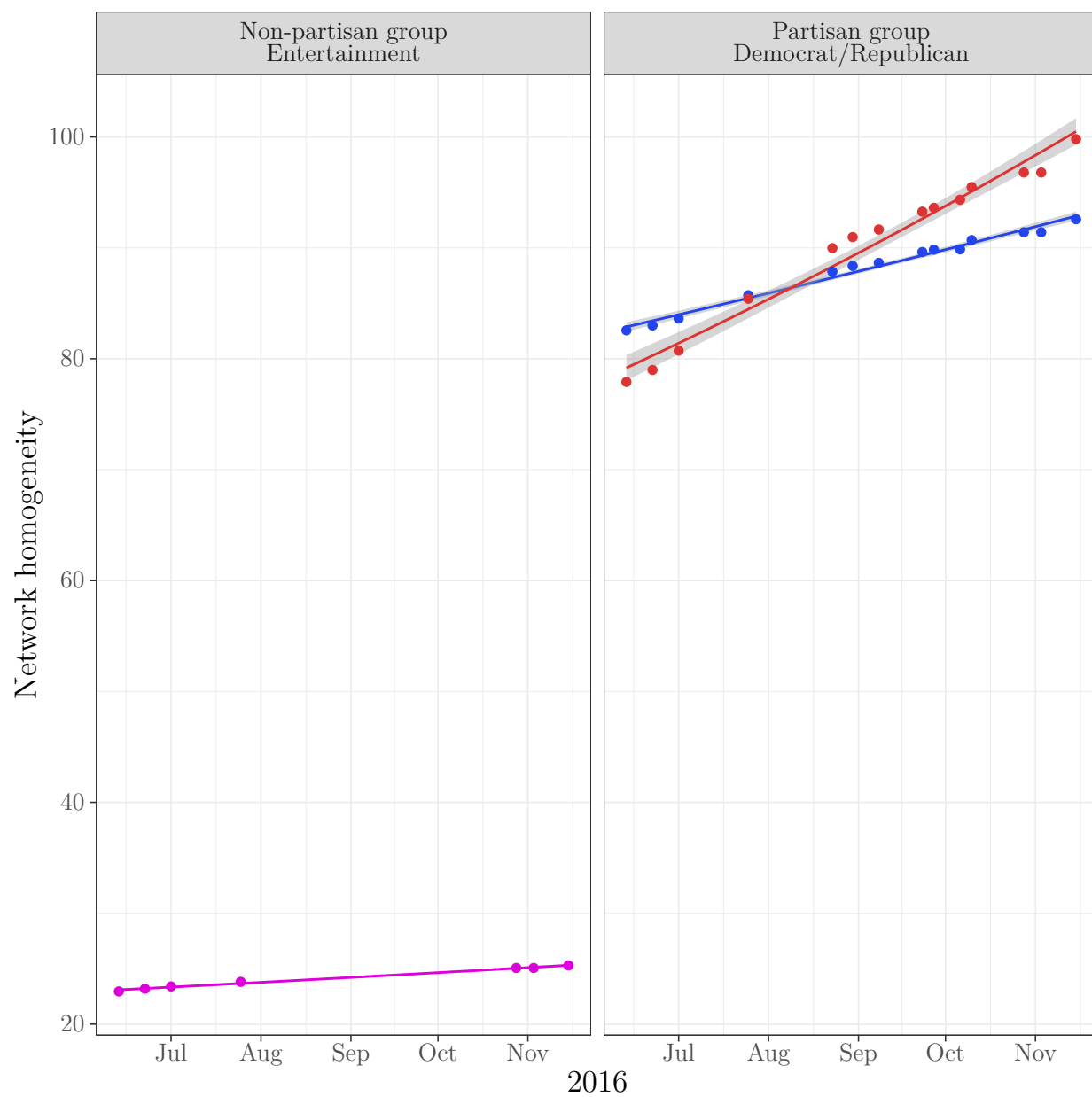


Figure 4.3: Change in number of weighted partisan follow-decisions by partisan groups

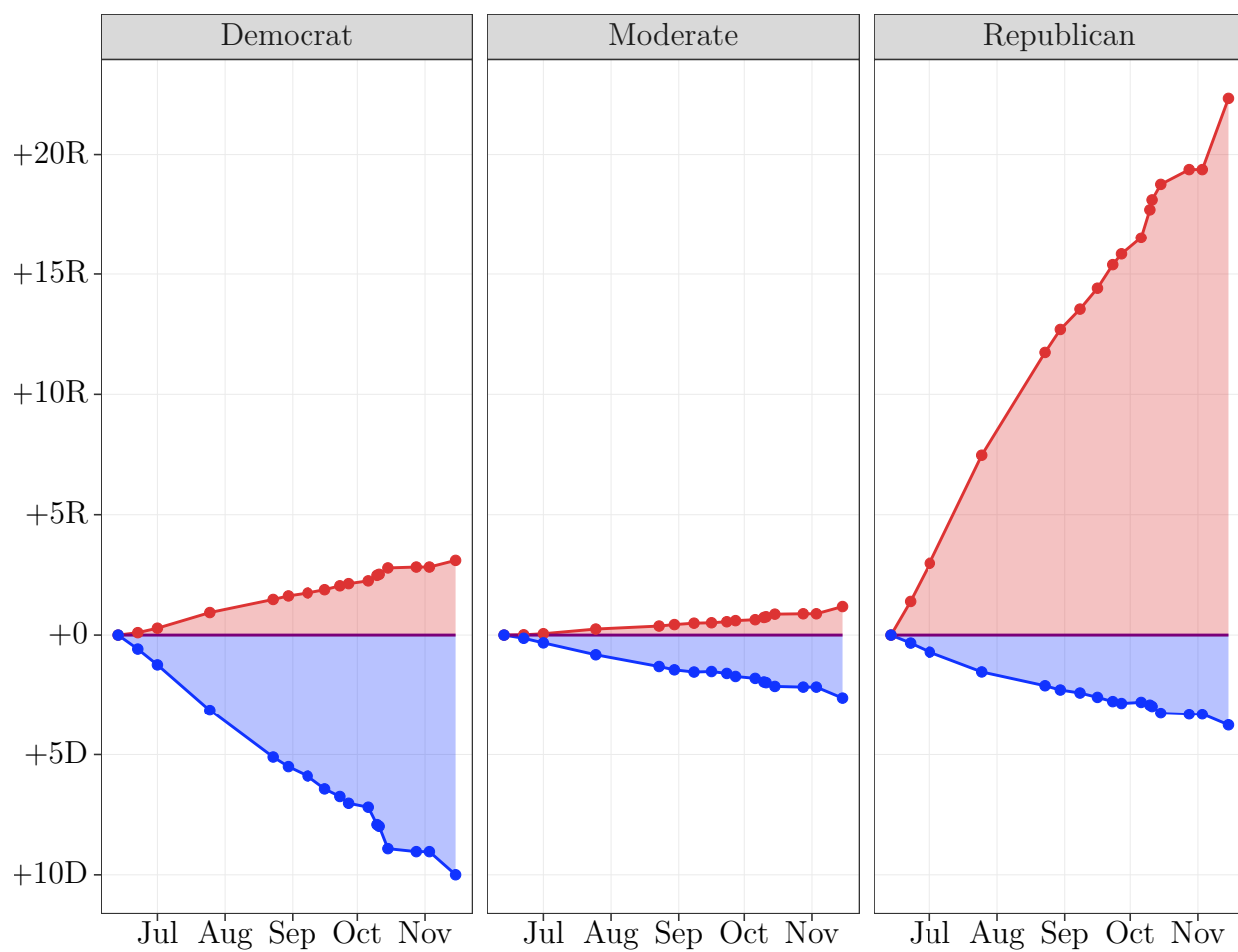
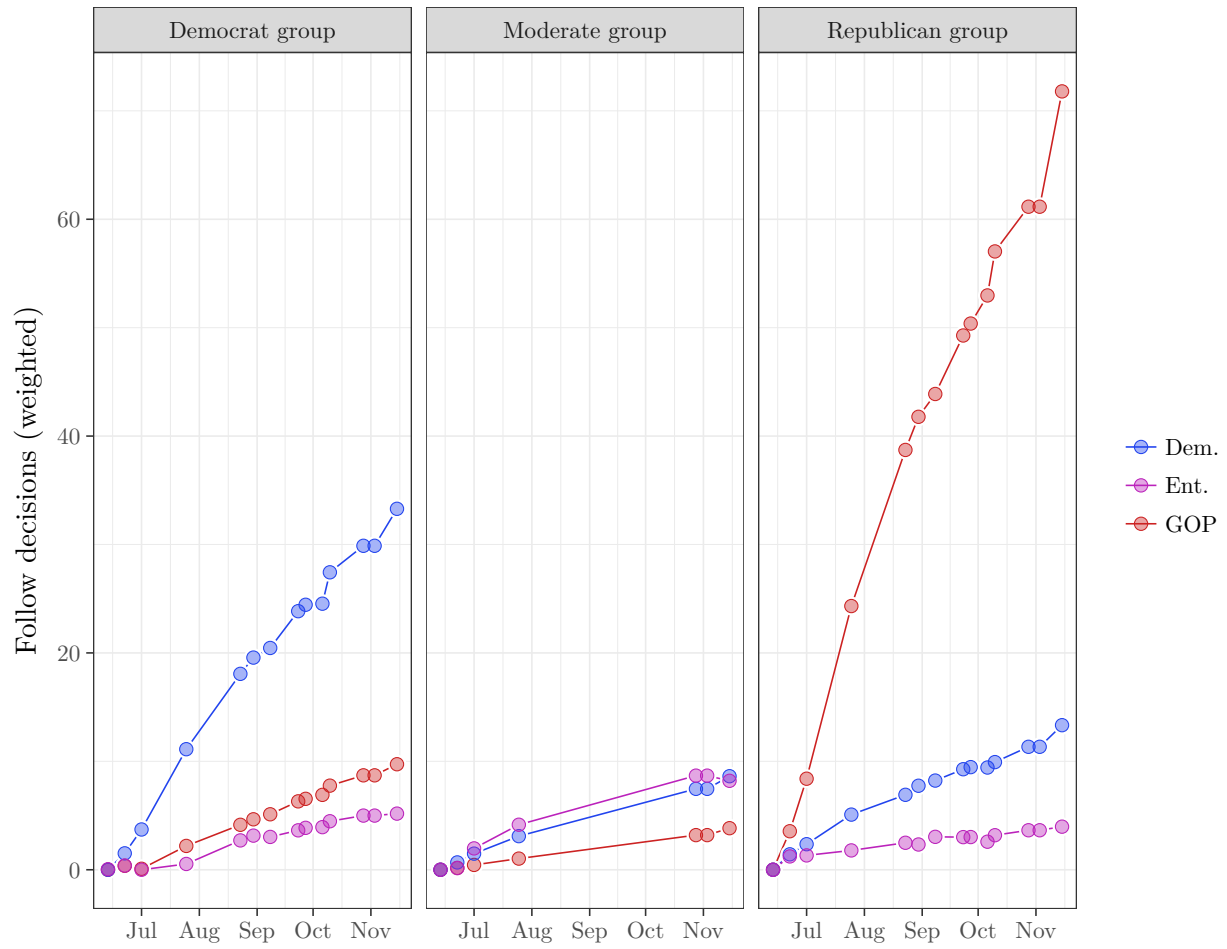


Figure 4.4: Change in number of partisan follow-decisions by all groups



## 4.6 Discussion

The present investigation is the first study of its kind to examine change in real-time behaviors of political polarization via network follow decisions during a politically-charged general election. The purpose of this study was to better understand change in network polarization on Twitter during a highly contested general election. Toward this end, social networks on Twitter were analyzed during the 2016 general election to provide a much needed perspective on political polarization in the new media environment. The sample consisted of followers randomly selected from well-known partisan and entertainment accounts. The data were collected by recording the entire friend network (all accounts followed by a user) of sampled users across seventeen time points—spanning from June, shortly after both major party candidates had become presumptive nominees, until November, shortly after election day. Data collection occurred during the final seven months leading up to the highly contentious 2016 general election.

### 4.6.1 Network polarization

Partisan users were hypothesized to engage in more politically homogeneous follow decisions than non-partisan users. Analyses therefore proceeded by examining the relationship between proximity to the election (coded as the number of weeks from the start of data collection) and within-subject change in network polarization. Overall, findings made good on a key assumption of the study—that the election would be noticed on Twitter. Results suggested that partisan users form highly partisan networks on Twitter, while moderate, or the less engaged, users mostly avoid politics. Over the full range of the data, findings suggest these patterns—partisan users with politically homogeneous networks and non-partisan users with mostly entertainment-focused follow decisions—held and, in the case of partisan users, intensified as the election got closer.

On the whole, these results present strong support for the hypothesis that change in

network homogeneity, or network polarization, increases with proximity to the election, especially among partisan users. The clearest evidence of this can be seen in Figure 4.3, which depicts the moving average of partisan follow decisions (or partisan composition of networks leading up to the election) from start to finish of data collection. As the figure plainly shows, partisan follow decisions of both partisan-democrats and partisan-republican users resulted in a clear pattern of network polarization as the election got closer, while the follow decisions of users in the entertainment/moderate group experienced only a minor up-tick in the number of partisan and non-partisan accounts that were added to networks.

Partisan users demonstrated a clear preference for homogeneous follow-decisions, but it should also be noted that the number of heterogeneous follow-decisions also increased over the course of the election for all groups. Thus, while the net result of follow-decisions was a larger total number of homogeneous compared to heterogeneous follow-decisions, this does not necessarily translate to an increase in the overall partisan homogeneity of user networks. For instance, if at the start of the study a user had followed 200 democrat elites and 50 republican elites, then adding two democrat elites for every one new republican elite during every week leading up to the election would actually result in a more politically heterogeneous user network. This result can be explained by the relative proportion of democrat follow-decisions at the start of the study (200 of 250 or .80) is greater than the proportion of democrat follow-decisions during the course of the study (2 of 3 or .67). One major limitation with this kind of proportional representation of homogeneity is that the values can only range between 0 and 1, which makes it especially difficult to detect increases in network homogeneity for users who already have a high proportion of homogeneous follow decisions. And because the goal of the present study was to understand change in *follow decisions* leading up to an election, the current study does not include analysis of change in the proportion of homogeneity.

One benefit to modeling the weighted number of follow-decisions as opposed to the proportion of homogeneous follow decisions is that they can reveal patterns found in the

total number of follow decisions. For instance, in the current study, the relative lack of partisan follow-decisions for users in the entertainment group provides additional evidence in support of the theory that proximity to an election has a greater influence on partisan users because non-partisan users simply continue to *tune out* of politics. In other words, the current study adds to the research suggesting that media choice reinforces existing gaps between the politically engaged/informed and disengaged/uninformed (e.g., [Arceneaux et al., 2013](#); [Bode et al., 2017](#); [Prior, 2005](#)). Evidence of this specifically comes from comparing the rate at which the non-partisan [entertainment] users added political accounts compared to the rate at which they added partisan [political] accounts. Post hoc regression analysis of the group-level means over time found change in non-partisan/entertainment follow decisions was not significantly different from change in partisan-democrat or partisan-republican follow decisions. In other words, the current study found users who tend to avoid politics do not suddenly become more interested in politics or politically active as a result of a highly-covered general election. Rather, users who prefer to “tune out” of politics continue to do so during contentious elections even if one of the leading candidates for president regularly made headlines for posts made on the same platform.

The gap in the effect of proximity to the election on partisan follow-decisions between partisan and non-partisan users may also explain seemingly contradictory findings from previous research on partisan selective exposure in social media. Although some research suggesting users cluster together according to partisanship (e.g, [Barberá, 2015](#); [Colleoni et al., 2014](#); [Himmelboim et al., 2013a](#)), other research suggests social media use positively associates with exposure to cross-cutting perspectives ([Lee et al., 2014](#); [Kim, 2011](#)). Implications from the current study offer one possible explanation for these conflicting findings. Results presented here suggest that whether social media use leads to partisan homogeneity may ultimately depend on the degree to which politics is salient. This would explain why partisan homogeneity is more pronounced on social media when users discuss political compared to non-political topics [Barberá et al. \(2015\)](#). In other words, social media does not inherently



increase exposure to diverse viewpoints nor does it inherently shelter users by creating self-reinforcing filter-bubbles. Rather, social media amplifies and reflects trends found in broader media environments.

The current study also demonstrates the importance and influence of social media on the U.S. political landscape. Between talk of “fake news” and a leading presidential candidate who frequently used his own social media account to express hostility toward “the media,” it is clear Twitter played a profound role during the 2016 presidential election (e.g., [Enli, 2017](#)), and the current study provides some of the first empirical evidence of it. Across all three groups—partisan republican, partisan democrat, and non-partisan [entertainment] users—partisan follow decisions increased over the course of the election. Partisan users not only followed more politically-oriented accounts in total, but they continued adding political accounts at higher rates than they did non-partisan (entertainment) accounts. And to the extent users in the sample demonstrated clear evidence of network polarization, the results presented here also suggests that decentralized [or fragmented] media environments like Twitter may even reinforce polarization behaviors among users in the mass public.

### **4.6.2 Limitations**

The current study has several limitations. Due to lack of controls and reliable measurement of exposure to politics on Twitter, it is impossible to make any definitive causal claims. Although unlikely, it is possible that follow decisions were driven by something other than the election or politics more generally that happened to line up with the partisan associations attributed to users in the sample. In addition, even if politics is attributed as a likely influence of follow decisions made during the study, evidence of polarization resulting from those decisions could still be more of a function of Twitter’s internal algorithms, e.g., algorithms responsible for making “people you may know” recommendations and other promotional features likely exploit an assumed preference for network homogeneity, than a function of considerate decisions made by users. Regardless of the mechanism, however, the implications

for network diversity and exposure to cross-cutting and democratic deliberation remain the same.

Another limitation of the current study is that it only describes a trend in follow decisions among three groups of users. The current study did not attempt to explain why individual follow-decisions were made or why they varied between individual users. Future research should investigate the specific motivations of both follow-decisions or unfollow-decisions. For instance, are follow decisions motivated by activity in the target account, such as an account posting statuses too frequently, or are they influenced by social endorsements and algorithm-based recommendation systems built-in to Twitter’s platform?

In summary, the current study aimed to provide the first in-depth explanation of change in network polarization during a highly-charged political election. Results confirmed the effect of partisan preferences on user behaviors on Twitter and provided evidence that those partisan preferences become amplified during political elections. Overall, the results provide undeniable evidence that partisan preferences play a major role in the organization of user networks and in the behaviors of social media users.

## Chapter 5

## Conclusion

## 5.1 Summary

### 5.1.1 Technological advancements

Recent advancements in technology have blurred the line between what is considered communication and what is considered data. Digital platforms increasingly—both in frequency and in scope—capture traces of communication and user behaviors, which has led to a growing number of companies making their data available to the public. Such developments may not completely alter the fundamental ways in which people interact, but they could represent a major shift in the way we approach the study of communication and media.

While recent advancements in media technologies have made it easier than ever before to access and use large amounts of publicly available social media data, the importation and wrangling of these sources of data remain as barriers to academic research. Addressing many of the practical obstacles which currently limit or prevent the use of social media data in academic research, Chapter 2 provided an introduction and brief overview of *rtweet*, an open-source R package designed to make collecting Twitter data and interacting with Twitter’s API a lot more approachable. In addition to highlighting a few of its major functions, Chapter 2 also situated and described the motivations for *rtweet*’s contributions.

### 5.1.2 Estimating partisanship and analyzing change

Following development of *rtweet*, Chapters 3 and 4 introduced, demonstrated, and then extended a new, network-based approach to the study of communication in a media environment. Using a network-based definition of partisanship, follow decisions of Twitter users sampled from well-known partisan and non-partisan accounts were analyzed in Chapter 3 (study II). User networks were then used to estimate the partisanship of thousands of accounts, many of which represent elected officials, media outlets, and popular media figures. Finally, Chapter 4 (study III) extended the network-based approach even further to analyze change network polarization among partisan and non-partisan users during the course of

the 2016 general election. Results showcase the wide range of uses for of a network-based approach, demonstrate the validity of network-based estimates of partisanship, and provide unequivocal evidence of highly-partisan user networks and an upward trend in polarization as proximity to the election increased.

### **5.1.3 Extending the network-based approach**

The potential for leveraging Twitter networks to measure various other social science constructs is nearly endless. This projects represents only a small glimpse of what is possible. Future communication and media research should extend this network-based approach and continue pushing the limits of what can be measured by collecting and analyzing these large digital sources of data. As such, this dissertation has introduced and demonstrated an accessible, systematic, and flexible approach—previously missing in communication and media research—to measuring communication phenomena in new media environments.

## **5.2 Communication, the hard science**

### **5.2.1 Computational social science**

The aim of this dissertation was to examine selective exposure in new media environments while also exploring new possibilities introduced by recent advancements in media technology. The primary goals of the project were to (1) make a theoretical contribution to the current understanding of partisan selective exposure and network polarization as situated in the contemporary media environment and (2) demonstrate an innovative and highly-scalable network-based approach to the study of communication and media. A side-effect of these contributions was development of computational social science skills and, subsequently, an open source package. Methodologists should continue to borrow from computer science methods to increase analysis capacity and maximize returns for all communication and media researchers. When possible, this kind of future work should also include contributions

designed to make software applications more approachable to a wide swath of scholars.

### 5.2.2 Future directions

Moving forward, researchers should leverage the network-based approach to estimating partisanship on Twitter to improve the accuracy of other definitions of partisanship, like those definitions mentioned in Study II. As other measures become more robust, they should be combined with network-based estimates and used to detect in real-time partisan behaviors online. Continuing in the tradition of also providing methodological contributions, future research should implement this kind of real-time detection system using machine learning methods, which represent the latest advancements in computer science methods for modeling prediction.

Machine learning refers to the use of automated algorithms and dynamic decision-models (similar to regression) in generating predictions ([Jordan & Mitchell, 2015](#)). As the algorithms experience and adapt to new data, prediction models and the accompanied predictions are refined over time—hence, the phrase “machine learning.” The goal in machine learning is to maximize the accuracy of predictions, which means theoretical contributions are rare. In applied settings, however, machine learning can be used to implement theoretically desirable models in applied settings. For instance, instead of displaying moving averages representing the frequency of hashtags on Twitter during political debates, media outlets could use machine learning and network-based definitions of partisanship to display the Twitter activity of partisan-democrat and partisan-republican users.

Ultimately, this line of research would make it possible to model, in real-time, both the content and audiences of mediated messages. With the rise of fake news websites designed to exploit and trick users, this research could contribute to systematic classifications of news media. Or, more broadly, this line of research could be used to provide the general public with dynamic and self-updating estimates of reliability, or credibility, for media sources. A media reliability index such as this could raise the standards for evidence used during con-

versations about public affairs between people in the general public, increase accountability of political campaigns and news media outlets, and improve transparency in the reporting of new information.

## Chapter 6

## Appendices



## 6.1 Study II

### 6.1.1 User groups

Figure 6.1 displays a visual depiction of the theorized user groups. To get a sense for how cleanly the results matched this theoretical model, picture this image juxtaposed on top of Figure 3.2.

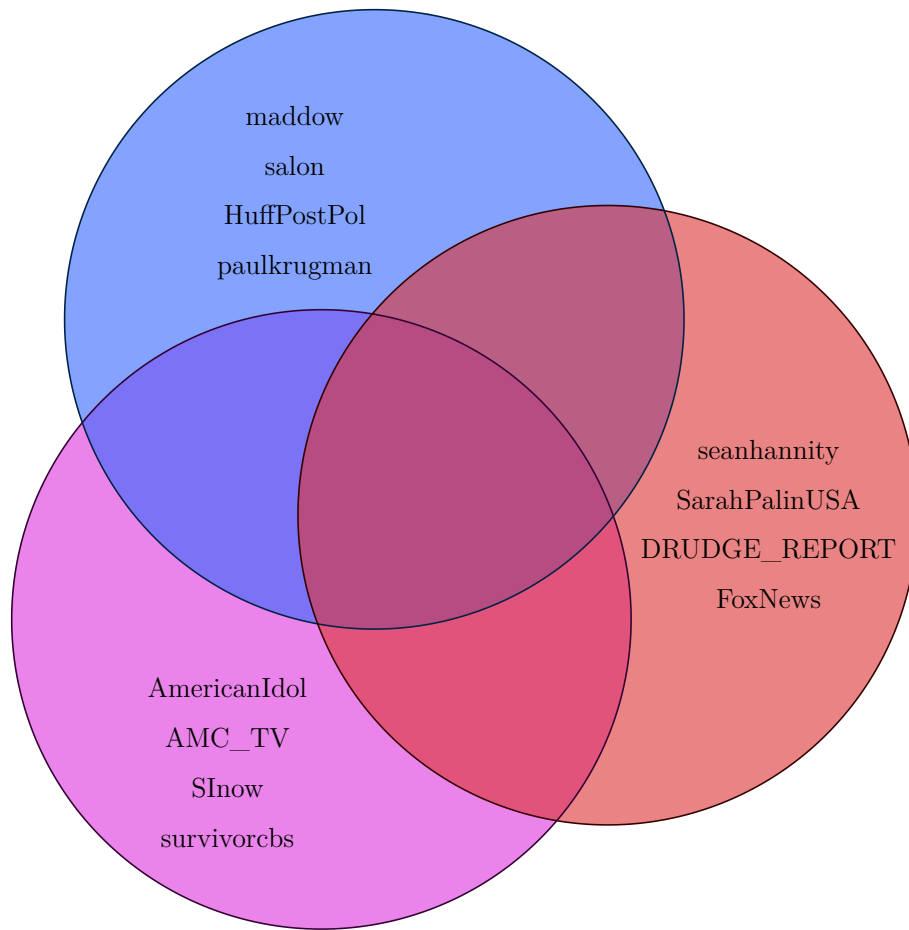


Figure 6.1: Twitter Source Accounts

### 6.1.2 Ordinary least squares model

Results from four ordinary least squares (OLS) regression models can be seen in Table 6.1. Each model, which are displayed along four columns, increases in complexity. The simplest model, Model 1, includes only user-level activity variables. Model 2 then adds a partisan variable—a dichotomous grouping variable indicating whether users were sampled from partisan (+1) or entertainment-centered (+0) accounts. In addition to user-level covariates and the partisan grouping variable, Model 2 includes a time variable, labeled “weeks,” which was calculated by dividing the number of days since the start of data collection (time 1 occurred on June 13, 2016) by 7. Finally, the most complex model, Model 4, includes all variables from the previous model and adds the interaction between the partisan grouping variable and the number of weeks since the start of data collection.

Using  $F$ -tests to sequentially compare models revealed that each model contributed significantly to explaining the variance of change in network polarization. Model 2 fit the data significantly better than Model 1,  $F(1) = 727.99, p < .05$ . Model 3 fit the data significantly better than Model 2,  $F(1) = 1,143.45, p < .05$ . And Model 4 fit the data significantly better than Model 3,  $F(1) = 244.43, p < .05$ . Although no directional hypotheses were made regarding the user-level activity variables, it is nevertheless interesting to note the magnitude with which change in network polarization was positively predicted by the number of friends a user had at time 1. This suggests, contrary to previous research suggesting social media increases exposure to diverse viewpoints, users who follow a lot of accounts are more likely to engage in follow decisions that result in more politically homogeneous networks as politics becomes more salient.

Overall, results presented in Table 6.1 suggest that differences in network homogeneity can be explained, at least in part, by the interaction of partisanship and proximity to the election. However, given the dichotomous nature of follow decisions—a data generating process which more closely resembles count data—it is possible the OLS estimates are biased, which could explain why the interaction between user partisanship and proximity to the

Table 6.1: OLS model estimates predicting change in partisan follow decisions

	Model 1	Model 2	Model 3	Model 4
(Intercept)	−9.33*** (1.27)	−28.62*** (1.45)	−45.78*** (1.52)	−29.36*** (1.85)
Account age	−2.49*** (.19)	−2.52*** (.19)	−2.62*** (.18)	−2.62*** (.18)
Statuses	.09 (.08)	.09 (.08)	.22** (.08)	.22** (.08)
Followers	−3.02*** (.16)	−2.94*** (.16)	−3.39*** (.16)	−3.39*** (.16)
Friends	8.31*** (.10)	8.16*** (.10)	8.15*** (.10)	8.14*** (.10)
Partisan		1.60*** (.06)	1.59*** (.06)	.30** (.10)
Weeks			28.11*** (.83)	3.63* (1.77)
Partisan*Weeks				1.94*** (.12)
R <sup>2</sup>	.13	.14	.16	.17
Adj. R <sup>2</sup>	.13	.14	.16	.17
Num. obs.	48,606	48,606	48,606	48,606
RMSE	87.17	86.53	85.53	85.32

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

election was no longer significant in the generalized linear model<sup>1</sup>.

---

<sup>1</sup> Linear models were estimated in base R ([R Core Team, 2017](#)). Generalized negative binomial models were provided via the MASS package to estimate change in discrete classifications—i.e., democrat coded as -1 and republican coded as 1 ([Venables & Ripley, 2002](#)).

### 6.1.3 Negative binomial models

The results for the generalized model can be found in Table 6.2. For the most part, results matched estimates from the OLS models. The notable exception was that the interaction term, which was significant in the OLS model, in Model 4 of Table 6.2 was not significant. Models were compared using a likelihood ratio test. The tests revealed improvements between Models 1-3, but Model 4 offered no such improvement over Model 3,  $-2LL(1) = -23,5416.51$ ,  $p = 0.845$ .

Table 6.2: Negative binomial estimates predicting change in network polarization

	Model 1	Model 2	Model 3	Model 4
(Intercept)	1.49*** (.03)	-.07* (.03)	-1.40*** (.03)	-1.41*** (.04)
Account age	-.09*** (.00)	-.10*** (.00)	-.07*** (.00)	-.07*** (.00)
Statuses	-.00* (.00)	-.00** (.00)	-.00 (.00)	-.00 (.00)
Followers	-.02*** (.00)	-.02*** (.00)	-.03*** (.00)	-.03*** (.00)
Friends	.13*** (.00)	.12*** (.00)	.12*** (.00)	.12*** (.00)
Partisan		.11*** (.00)	.11*** (.00)	.11*** (.00)
Weeks			1.52*** (.02)	1.53*** (.04)
Partisan*Weeks				-.00 (.00)
AIC	246,360.44	241,339.77	235,432.55	235,434.51
BIC	246,413.19	241,401.31	235,502.88	235,513.64
Log Likelihood	-123,174.22	-120662.88	-117708.28	-117,708.26
Deviance	49,067.49	48,766.41	48,482.01	48,481.67
Num. obs.	48,606	48,606	48,606	48,606

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

## 6.2 Study III

### 6.2.1 Missing data

As shown in the bold values in Table 6.3, a streak of missingness occurred among the entertainment group in the middle waves of data collection. This was due to a bug in the code that was not immediately noticed.

Table 6.3: Number of observations per time point by group

Wave	Sample groups			
	Democrat	Entertainment	Republican	
1	1000	1000	1000	3000
2	998	991	992	2981
3	991	987	984	2962
4	981	985	971	2937
5	973	<b>904</b>	962	2839
6	971	<b>899</b>	957	2827
7	969	<b>901</b>	957	2827
8	971	<b>899</b>	953	2823
9	965	<b>892</b>	949	2806
10	964	<b>892</b>	949	2805
11	963	<b>890</b>	947	2800
12	963	<b>890</b>	942	2795
13	964	<b>888</b>	935	2787
14	963	953	934	2850
15	963	958	937	2858
16	963	958	937	2858
17	960	955	936	2851

### 6.2.2 Source accounts

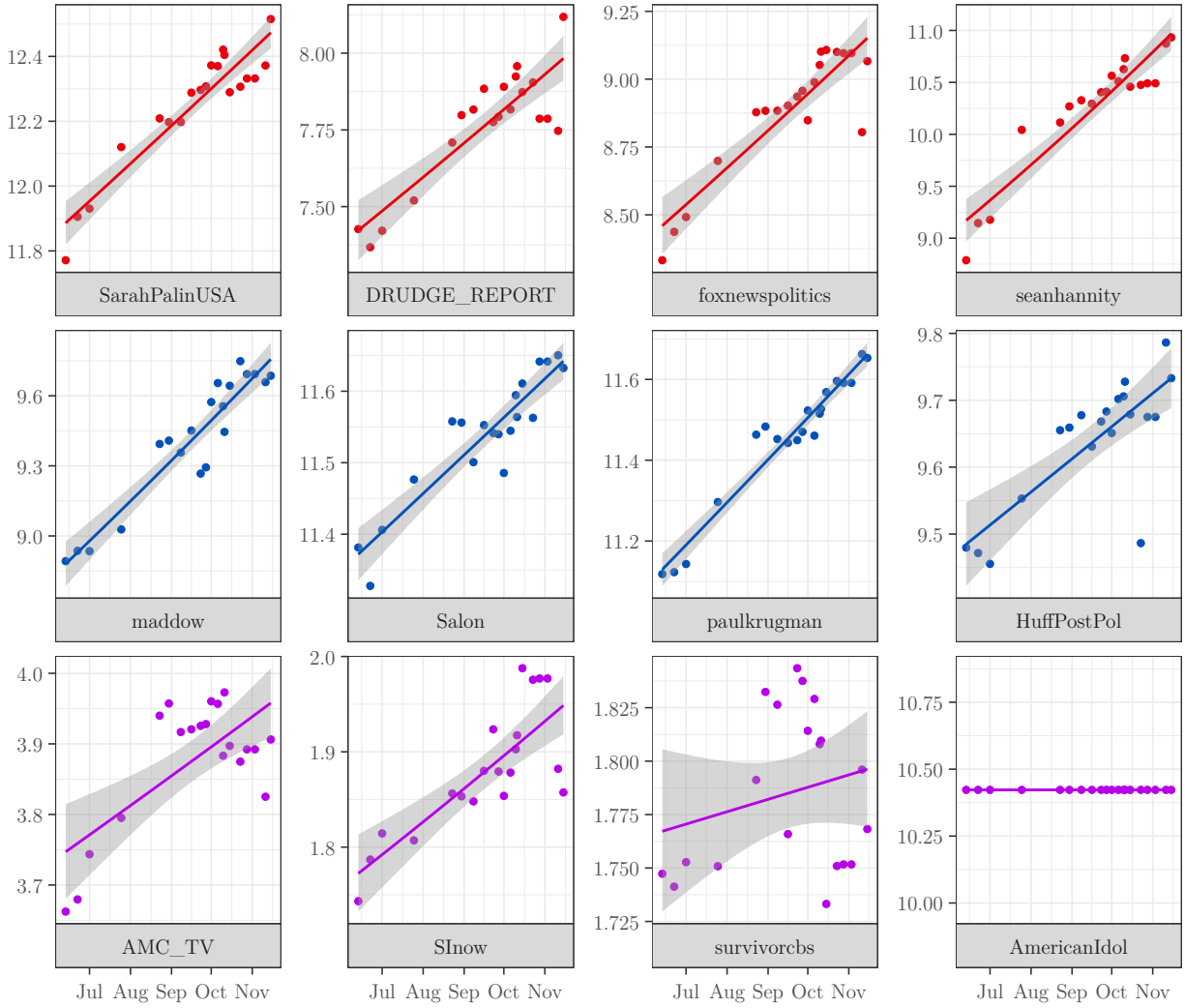
Users were originally sampled from twelve source accounts—four chosen to represent each of the partisan and non-partisan groups. Eigenvalue decomposition confirmed the assumed group assignments for all accounts.<sup>2</sup> However, for those interested in how change in partisan follow decisions broke down by source account, a visual depiction is provided in Figure 6.2.<sup>3</sup>

---

<sup>2</sup> Due to filtering rules and random chance, only one **AmericanIdol** follower was selected, which means that even though follow decisions of the single user correlated with follow decisions made by users sampled from other entertainment source accounts, there was not enough evidence to distinguish **AmericanIdol** from partisan accounts. However, given the difference between the average loading for the other entertainment accounts on the same component ( $M = .74$ ) compared to the average loading for the other two components ( $M = .04$ ), the user was still included as part of the entertainment group.

<sup>3</sup> See the previous footnote for an explanation of the sample size ( $n = 1$ ) for the **AmericanIdol** facet, which explains its unique pattern compared to all other source accounts.

Figure 6.2: Change in partisan scores by account





# References

- Abramowitz, A. I. & Saunders, K. L. (1998). Ideological realignment in the US electorate. *The Journal of Politics*, 60(3), 634–652.
- Ansolabehere, S. D., Hirano, S., Hansen, J. M., & Snyder Jr, J. M. (2010). Primary elections and partisan polarization in the US congress. *Quarterly Journal of Political Science*, 5, 169—191.
- Arceneaux, K., Johnson, M., & Cryderman, J. (2013). Communication, persuasion, and the conditioning value of selective exposure: Like minds may unite and divide but they mostly tune out. *Political Communication*, 30(2), 213–231.
- Arceneaux, K., Johnson, M., & Murphy, C. (2012). Polarized political communication, oppositional media hostility, and selective exposure. *The Journal of Politics*, 74(1), 174–186.
- Baldassarri, D. & Gelman, A. (2008). Partisans without constraint: Political polarization and trends in American public opinion. *American Journal of Sociology*, 114(2), 408–446.
- Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1), 76–91.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542.

- Barberá, P. & Rivero, G. (2014). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, 33(6), 712–729.
- Bartels, L. M. (2000). Partisanship and voting behavior, 1952-1996. *American Journal of Political Science*, (pp. 35–50).
- Bennett, W. L. & Iyengar, S. (2008). A new era of minimal effects? the changing foundations of political communication. *Journal of Communication*, 58(4), 707–731.
- Bernaards, C. A. & Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65, 676–696.
- Bode, L. (2016). Pruning the news feed: Unfriending and unfollowing political content on social media. *Research & Politics*, 3(3), 1–8.
- Bode, L., Vraga, E. K., & Troller-Renfree, S. (2017). Skipping politics: Measuring avoidance of political content in social media. *Research & Politics*, 4(2), 2053168017702990.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298.
- Breen, J. O. (2012). Mining twitter for airline consumer sentiment. *Practical text mining and statistical analysis for non-structured text data applications*, 133.
- Carson, J. L., Crespin, M. H., Finocchiaro, C. J., & Rohde, D. W. (2007). Redistricting and party polarization in the US house of representatives. *American Politics Research*, 35(6), 878–904.
- Clay, R., Barber, J. M., & Shook, N. J. (2013). Techniques for measuring selective exposure: A critical review. *Communication Methods and Measures*, 7(3-4), 147–171.

- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.
- Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. *ICWSM*, 133, 89–96.
- DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American’s social attitudes become more polarized? *American journal of Sociology*, 102(3), 690–755.
- Enli, G. (2017). Twitter as arena for the authentic outsider: Exploring the social media campaigns of trump and clinton in the 2016 US presidential election. *European Journal of Communication*, 32(1), 50–61.
- Fiorina, M. P. & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11, 563–588.
- Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, 35(4), 469–492.
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users<sup>1</sup>. *Journal of Computer-Mediated Communication*, 14(2), 265–285.
- Garrett, R. K. & Stroud, N. J. (2014). Partisan paths to exposure diversity: Differences in pro-and counterattitudinal news consumption. *Journal of Communication*, 64(4), 680–701.
- Gelman, A. et al. (2005). Analysis of variance— why it is more important than ever. *The Annals of Statistics*, 33(1), 1–53.
- Gentry, J. (2013). *twitteR: R based Twitter client*. Comprehensive R Archive Network.

- Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3), 319–336.
- Glaeser, E. L. & Sunstein, C. R. (2013). *Why does balanced news produce unbalanced views?* Technical report, National Bureau of Economic Research.
- Gleason, S. (2010). Harnessing social media: News outlets are assigning staffers to focus on networking. *American Journalism Review*, 32(1), 6–8.
- Gottfried, J. & Shearer, E. (2017). News use across social media platforms. *Pew Research Center*.
- Gramlich, J. (2016). America's political divisions in 5 charts. *Pew Research Center*.
- Harris, J. K. (2015). Diabetes topics associated with engagement on twitter. *Preventing chronic disease*, 12.
- Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, 67(1), 232–238.
- Hetherington, M. J. (2001). Resurgent mass partisanship: The role of elite polarization. In *American Political Science Association*, volume 95 (pp. 619–631).: Cambridge Univ Press.
- Hill, S. J. & Tausanovitch, C. (2015). A disconnect in representation? comparison of trends in congressional and public polarization. *The Journal of Politics*, 77(4), 1058–1075.
- Himelboim, I. (2014). Political television hosts on Twitter: Examining patterns of interconnectivity and self-exposure in Twitter political talk networks. *Journal of Broadcasting & Electronic Media*, 58(1), 76–96.

- Himmelboim, I., McCreery, S., & Smith, M. (2013a). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2), 40–60.
- Himmelboim, I., Smith, M., & Shneiderman, B. (2013b). Tweeting apart: Applying network analysis to detect selective exposure clusters in Twitter. *Communication Methods and Measures*, 7(3-4), 195–223.
- Hmielowski, J. D., Beam, M. A., & Hutchens, M. J. (2015). Structural changes in media and attitude polarization: Examining the contributions of TV news before and after the Telecommunications Act of 1996. *International Journal of Public Opinion Research*, (pp. edv012).
- Holbert, R. L., Garrett, R. K., & Gleason, L. S. (2010). A new era of minimal effects? a response to Bennett and Iyengar. *Journal of Communication*, 60(1), 15–34.
- Holbert, R. L., Hmielowski, J. D., & Weeks, B. E. (2012). Clarifying relationships between ideology and ideologically oriented cable tv news use a case of suppression. *Communication Research*, 39(2), 194–216.
- Hollander, B. A. (2008). Tuning out or tuning elsewhere? partisanship, polarization, and media migration from 1998 to 2006. *Journalism & Mass Communication Quarterly*, 85(1), 23–40.
- Hong, S. (2012). Online news on Twitter: Newspapers' social media adoption and their online readership. *Information Economics and Policy*, 24(1), 69–74.
- Iyengar, S., Hahn, K. S., Krosnick, J. A., & Walker, J. (2008). Selective exposure to campaign communication: The role of anticipated agreement and issue public membership. *The Journal of Politics*, 70(01), 186–200.

- Jang, S. M. & Pasek, J. (2015). Assessing the carrying capacity of Twitter and online news. *Mass Communication and Society*, 18(5), 577–598.
- Jones, D. A. (2002). The polarizing effect of new media messages. *International Journal of Public Opinion Research*, 14(2), 158–174.
- Jordan, M. I. & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2013). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, Forthcoming; Yale Law School, Public Law Working Paper No. 307*.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kearney, M. W. (2016). *rtweet: Collecting Twitter data*. Comprehensive R Archive Network.
- Kearney, M. W. (2017). A network-based approach to estimating partisanship. *Dissertation*, (pp. 15–30).
- Kenski, H. C. (1980). Running with and from the PAC. *Ariz. L. Rev.*, 22, 627.
- Kim, Y. (2011). The contribution of social network sites to exposure to political difference: The relationships among SNSs, online political messaging, and exposure to cross-cutting perspectives. *Computers in Human Behavior*, 27(2), 971–977.
- Knobloch-Westerwick, S. & Johnson, B. K. (2014). Selective exposure for better or worse: Its mediating role for online news’ impact on political participation. *Journal of Computer-Mediated Communication*, 19(2), 184–196.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480.
- Larsson, A. O. & Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, 14(5), 729–747.

- Lau, R. R., Andersen, D. J., Ditonto, T. M., Kleinberg, M. S., & Redlawsk, D. P. (2016). Effect of media environment diversity and advertising tone on information search, selective exposure, and affective polarization. *Political Behavior*, (pp. 1–25).
- Layman, G. C. & Carsey, T. M. (2002). Party polarization and "conflict extension" in the American electorate. *American Journal of Political Science*, (pp. 786–802).
- Lee, F. L. (2016). Impact of social media on opinion polarization in varying times. *Communication and the Public*, 1(1), 56–71.
- Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, 64(4), 702–722.
- Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. University of Chicago Press.
- Levendusky, M. S. (2013). Why do partisan media polarize viewers? *American Journal of Political Science*, 57(3), 611–623.
- Levendusky, M. S. & Malhotra, N. (2016). (mis) perceptions of partisan polarization in the American public. *Public Opinion Quarterly*, 80(S1), 378–391.
- Lodge, M. & Taber, C. (2000). Three steps toward a theory of motivated political reasoning. *Elements of reason: Cognition, choice, and the bounds of rationality*, (pp. 183–213).
- Mancini, P. (2013). Media fragmentation, party system, and democracy. *The International Journal of Press/Politics*, 18(1), 43–60.
- Mann, T. E. (2007). Polarizing the house of representatives: How much does gerrymandering matter? *Red and blue nation*, (pp. 263–283).
- McCarty, N., Poole, K. T., & Rosenthal, H. (2009). Does gerrymandering cause polarization? *American Journal of Political Science*, 53(3), 666–680.

- Messing, S. & Westwood, S. J. (2012). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, (pp. 0093650212466406).
- Mitchell, A., Gottfried, J., Kiley, J., & Matsa, K. E. (2014). Political polarization & media habits. *Pew Research Center*.
- Mouffe, C. (2005). *The return of the political*, volume 8. Verso.
- Nyhan, B. & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.
- Pew Research Center (2017). Three technology revolutions. *Pew Research Center*. Retrieved from <http://www.pewinternet.org/three-technology-revolutions/>.
- Prior, M. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, 49(3), 577–592.
- Prior, M. (2009a). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, 73(1), 130–143.
- Prior, M. (2009b). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, 71(3), 893–908.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101–127.
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W. (2016). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois. R package version 1.6.12.



- Robinson, D. (2016). Two people write trump's tweets: He writes the angrier ones. *Washington Post*.
- Shin, J., Jian, L., Driscoll, K., & Bar, F. (2016). Political rumoring on Twitter during the 2012 us presidential election: Rumor diffusion and correction. *New Media & Society*, (pp. 1–22).
- Smith, S. M., Fabrigar, L. R., & Norris, M. E. (2008). Reflecting on six decades of selective exposure research: Progress, challenges, and opportunities. *Social and Personality Psychology Compass*, 2(1), 464–493.
- Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30(3), 341–366.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556–576.
- Taber, C. S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769.
- Theriault, S. M. & Rohde, D. W. (2011). The Gingrich senators and party polarization in the US senate. *The Journal of Politics*, 73(4), 1011–1024.
- Tufekci, Z. & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), 363–379.
- Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer, fourth edition. ISBN 0-387-95457-0.
- Vidal, L., Ares, G., Machín, L., & Jaeger, S. R. (2015). Using twitter data for food-related consumer research: A case study on what people say when tweeting about different eating situations. *Food Quality and Preference*, 45, 58–69.

- Warner, B. R., McGowen, S. T., & Hawthorne, J. (2012). Limbaugh's social media nightmare: Facebook and Twitter as spaces for political action. *Journal of Radio & Audio Media*, 19(2), 257–275.
- Warner, B. R. & McKinney, M. S. (2013). To unite and divide: The polarizing effect of presidential debates. *Communication Studies*, 64(5), 508–527.
- Westen, D. (2008). *Political brain: The role of emotion in deciding the fate of the nation*. PublicAffairs.
- Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, 10(2), 145–158.
- Wicks, R. H., Wicks, J. L., & Morimoto, S. A. (2014). Partisan media selective exposure during the 2012 presidential election. *American Behavioral Scientist*, 58(9), 1131–1143.
- Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, 7(3), 203–220.
- Yardi, S., Romero, D., Schoenebeck, G., et al. (2009). Detecting spam in a Twitter network. *First Monday*, 15(1).
- YouGov (2013). Sept 27-28 poll on reading. *Research*.
- Younis, E. M. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: An empirical study. *International Journal of Computer Applications*, 112(5).